

## From red peril to red tape

**Tight US customs controls on microbiological materials could threaten foreign research collaborations. Especially at risk are attempts to integrate Russia's former bioweapons scientists into the mainstream.**

**W**hen Thomas Butler of Texas Tech University in Lubbock was arrested in January, a clear message was sent out to US microbiologists working with foreign collaborators: stick to the letter of the law — or else. Butler was working to develop new antibiotics to treat plague. But the project is in disarray after he was accused of various offences, including smuggling vials containing the plague bacterium, *Yersinia pestis*, collected in Tanzania, into the United States on commercial flights.

Butler's case has yet to come to trial. But other researchers who work on similar pathogens admit privately that they have been tempted to bend the rules that Butler is accused of flouting. Thanks to strict customs controls introduced under the ongoing 'war on terrorism', it has become all but impossible to bring samples of pathogenic bacteria and viruses into the United States through official channels.

Those trying to initiate collaborations with former members of the Soviet bioweapons establishment are especially frustrated. They must fight customs authorities in both countries to get even basic equipment sent to Russia, and wrestle with immigration officials to invite collaborators back to their labs. Their efforts are hampered by mutual mistrust between former Cold War opponents: while US officials are worried by tales of endemic Russian corruption, their Russian counterparts seem to view each visiting Western scientist as a potential spy (see page 678).

But these collaborations stand to benefit both sides. The US government is currently investing enormous sums in research to counter the threat of bioterrorism. For the scientists charged with this task, it makes sense to reach out to the labs that once served as the research powerhouses of the world's largest biowarfare programme. On the other side, Russia's pressing problems with epidemics of infectious diseases, including AIDS and tuberculosis, cry out for the skills of its

former military microbiologists to be turned to peaceful purposes.

Things must change if these collaborations are to continue. Russian officials must learn to welcome Western scientists into their labs. And the United States needs to review its sometimes overzealous customs and immigration controls. In the wake of 11 September 2001, and the anthrax attacks that followed, clamping down on the import of hazardous biological material was sensible. But risks need to be put into perspective. It is arguable that the current expansion of domestic US labs working on biodefence poses a greater threat of proliferation than the transfer of samples between the United States and Russia.

If Western officials are serious about countering bioweapons proliferation, they should also start thinking more broadly. Building capacity in microbiology outside the secret network of labs that housed the Soviet Union's biowarfare programme is important, to help integrate the labs' denizens into the broader scientific community of the former Soviet states. One positive example is a European Union-funded project under which scientists in Oslo and London helped to analyse data on samples of fleas and rodents collected by the Anti-Plague Research Institute in Almaty, Kazakhstan, over some 50 years. The project has improved understanding of the transmission of plague in central Asia.

There are a few encouraging signs. US government officials are reportedly talking to the Russian Academy of Sciences in the hope of getting it interested in working with the former bioweapons labs. But integrating former bioweapons scientists into the international scientific community seems to be a low priority for politicians on both sides. Until legislators realize the sizeable benefits to both Russia and the rest of the world, scientists working on such collaborations will be faced with an unenviable choice: let their research stagnate, or risk breaking the law. ■

## Debate, what debate?

**The UK government is squandering the chance to canvass public opinion on one of the hottest controversies in science.**

**H**ow do you condense the views of the public on genetically modified (GM) crops into a single document? One innovative, if experimental, solution is to try to spark public debate among community groups across the country. In Britain, this effort got under way last week. It is supposed to inform the decision, to be made in the autumn, on whether to allow commercial planting of the crops.

Six set-piece debates are being held in cities around Britain in the first two weeks of June. These are meant to act as springboards for discussions at local level. Questionnaires and recordings from the events should generate more detailed insights into public thinking than opinion polls can provide — helping to reveal how people have come to hold the views that they express (see page 672).

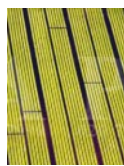
It's a laudable idea. But one problem has become painfully clear: most of the public don't know they are invited, for reasons that aren't hard to discern. Over the past three years, the Netherlands (population 16 million) and New Zealand (population 4 million) have conducted similar programmes to assess opinion on genetically modified crops,

each investing some four times the sum allotted in Britain (population 60 million). This penny-pinching has restricted advertising, and turnouts at the first debates have been limited to a few hundred, with the majority already having a vested interest in the subject.

Members of the panel organizing the debate, which includes representatives from industry and academia, as well as experts on public consultation, claim that they were given too little time to set up the debate. The panel's first meeting was last September; the government expects the results to be complete by mid-July.

This sorry state of affairs will have two consequences. The opportunity to test a consultation process that could be applied to many other scientific controversies — from the alleged environmental hazards posed by nanotechnology to the ethics of embryonic stem-cell research — could be squandered. Worse, negative media coverage may leave the British people to assume that the government has already made up its mind on transgenic agriculture, and simply isn't interested in their views. ■





**Straight talking**  
UK government kicks off public discussions on transgenic crops  
**p672**



**Running dry**  
Weak monsoon may not be enough for a thirsty India  
**p673**



**Down the drain**  
Italian neutrino lab closed over fears of toxic chemical leak  
**p675**



**Sputnik for sale**  
Russian satellite goes up for grabs at car dealership  
**p677**

# Astronomers leap to defence of extra seconds in time debate

**Geoff Brumfiel, Washington**

What time is it? For air-traffic controllers, astronomers, computer network managers and others in search of the *real* answer to this everyday question, a battle royal is set to erupt.

Experts are trying to agree on whether the long-standing practice of inserting occasional 'leap seconds' into coordinated universal time — the standard time reference used by clocks around the world — should be abandoned. The body that will decide how to resolve the issue, the International Telecommunication Union (ITU), is now asking astronomers for their input.

Since the practice began in 1972, 32 leap seconds have been added to universal time, mainly to keep it in synch with the rotation of the Earth as it slows down under the gentle drag of the Moon's gravitational field.

Astronomers want to continue adding leap seconds because keeping universal time in line with the Earth's rotation means that telescopes can find stars and galaxies in their rightful places in the sky. But no one navigates by the stars any more, and adding leap seconds creates extra work all round.

Most of those involved want to quit the practice, and get universal time back in alignment with the other main reference for timekeepers — atomic time, as assiduously maintained by the Bureau International des Poids et Mesures in Paris. If the change is implemented, "the old meaning of time will become ambiguous," warns Steve Allen, an astronomer at the Lick Observatory in Santa Cruz, California.

But William Klepczynski of Global Timing Services, which consults for the US Federal Aviation Administration, fears that keeping leap seconds will lead to trouble, perhaps even big trouble — such as a plane crash. "There have been no accidents yet, but I think the potential is there," he contends.

Confusion caused by the divergence of universal time and atomic time has been further exacerbated by the architects of the Global Positioning System (GPS). The GPS began on universal time in 1980 but wasn't adjusted by adding any of the 13 leap seconds since then because its carefully synchronized



**Time for a change?** The issue of leap seconds has split researchers who need an accurate time standard.

satellite network can't cope with them.

As a result, Klepczynski says, gaps of several seconds' duration are opening up between the world's navigation systems and the universal time used by pilots and air-traffic controllers. These can confuse sensitive aircraft navigation systems, and the only way to avoid disaster, he says, is to detach universal time from the Earth's rotation and make it the standard for all navigation.

But separating time from rotation would mean that telescopes would no longer function properly, according to Patrick Wallace, an astronomer at the Rutherford Appleton Laboratory in Oxford, UK. Most telescopes with a viewfinder wouldn't be immediately affected, but instruments that track satellites and other moving objects could feel the effects within a few years, Wallace warns.

Allen says that it could cost between US\$10,000 and US\$100,000 for each telescope to correct the problem, and that corrections will be tricky to implement.

Then there is the problem that, without the leap seconds, day will turn into night.

"In principle, you could eventually find that you're having lunch when it's getting dark," Wallace says.

Compromise solutions don't seem to please anyone. Wallace says that an idea to change the length of the second is "completely lunatic". A more prosaic plan to replace the leap second with a leap hour, to be inserted in a few centuries' time, is condemned by Klepczynski as "inelegant".

The decision about whether to drop the leap second is being studied by a working group on time signals set up by the ITU's Radiocommunication sector. The group's chairman, Ronald Beard of the Naval Research Laboratory in Washington DC, says he needs more input from people who use universal time. "Many people don't even know that leap seconds exist," he complains.

Beard's working group hopes to come up with a plan in time for the ITU's next World Radiocommunication Conference, probably in 2006 (see page 675). But it could take years for the ITU to finally agree what time it is, he predicts.

# Public input sought on transgenic farming

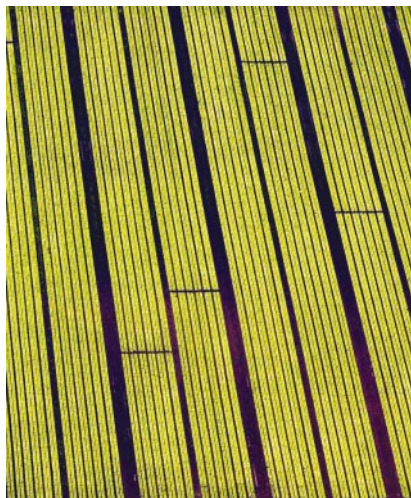
Claire Tilstone, London

A fresh method for assessing public opinion on scientific matters is being tested in Britain this month. But observers of the process, which aims to assess opinion about genetically modified crops through a series of public debates, say it may need more money and better planning if it is to succeed.

The debates are part of the government's attempts to assess opinion ahead of this autumn's decision on whether to allow commercial planting of herbicide-resistant crops. Some conventional methods are being employed — standard social-research protocols, for example, are being used in focus groups. But the panel of academics and industry representatives charged with overseeing the process has also opted for a more novel approach — six public debates, held this month in towns around Britain, which are designed to spark off many smaller debates among community organizations.

When *Nature* attended the second debate, held in Swansea in Wales on 5 June, things got off to a shaky start. Those who arrived for the advertised 18:00 start found they had missed the introductory video, which had begun half an hour before. And attendees were split into groups and presented with a series of fact sheets, rather than being allowed to question a panel of experts. "There isn't anyone here who can tell us the facts," said one participant.

The ensuing hour-long debate proved to be a bumpy ride. A local organic farmer, who was



**Strands of opinion:** transgenic crops are a source of continuing public debate in Britain.

strongly against transgenic crops, had reams of documents backing up points she made. Her assertive approach seemed to make it difficult for those with less experience of the topic to get involved. At the end of the debate, none of the eight members of *Nature's* table had changed their feelings about transgenic crops.

The events are meant to encourage people to set up their own smaller debates among friends, colleagues or community groups, and supporting materials such as a CD-ROM and video are available. The results of these smaller

debates will be collected through a questionnaire that asks people to what extent they agree with certain statements about genetically modified crops and invites them to give their views. They can also complete this questionnaire on the 'GM Nation?' website. The government hopes by this to develop a more sophisticated picture of public opinion than surveys and focus groups can provide.

But those involved admit that the debates may not be reaching people who are undecided about transgenic crops but wish to know more. The Swansea meeting, like the opening meeting in Birmingham on 3 June, attracted just a few hundred people.

The low attendance, and lack of interest from people new to the topic, may be due to poor advertising. The government originally allocated the debates just £250,000 (US\$410,000) in funding, upped to £500,000 after complaints from the panel of organizers. Similar debates held in New Zealand and the Netherlands over the past three years received around £2 million in government funding.

"We have been severely constrained by money," says Gary Kass, an adviser at the Parliamentary Office of Science and Technology, and a member of panel. But ministers insist that they will consider the results of the debates alongside the findings of an ongoing review of scientific studies of transgenic crops, together with the results of a farm-scale trial of such crops, which began in 2000. ■

♦ [www.gmpublicdebate.org](http://www.gmpublicdebate.org)

## All change as Argentina's science leader keeps his job

Carol Marzuola, Caracas

Argentina hasn't enjoyed much continuity of late. So hard-pressed scientists are relieved that newly elected president Néstor Kirchner plans to leave the country's top scientific administrator in place.

Kirchner's administration will retain Eduardo Charreau as the president of the National Council for Science and Technology (CONICET), the country's main science agency. "It's the first time in history that CONICET authorities won't change with a new president," says Charreau, a molecular endocrinologist and former director of the Institute of Biology and Experimental Medicine in Buenos Aires.

Tulio Del Bono, a political ally of the president, has been appointed secretary of science and technology, overseeing CONICET. Del Bono, an engineer and former rector of the National University of San Juan, helped Kirchner on science and technology aspects of his campaign — most notably, a bold pledge to increase total

spending on research and development from 0.4% to 1.0% of economic output by 2006.

This year, the Argentinian government will spend about \$400 million on supporting an estimated 21,000 researchers, including about 3,500 scientists at 116 CONICET laboratories.

CONICET's budget for 2003 is up 30% in Argentinian pesos to a total of \$80 million — still little more than half the \$150 million it received two years ago, before the drastic devaluation of the peso. But Charreau admits that much more is needed to rescue research from the impact of the devaluation and associated high inflation. "Frankly," he says, "the system is beat up and needs an urgent budgetary reactivation."

Charreau says his biggest challenge has been simply to "keep the doors of its institutions open" — for which Argentinian scientists thank him and departing science and technology secretary Julio Luna.

Del Bono says his main aim will be to attract more private money into research,

to address social and economic problems. "Our emphasis and specific instructions are to fortify science and technology and to ensure that production coming out of this system can be rapidly applied," he says.

Top research priorities will be health, agriculture and livestock, and sustainable development of natural resources, he says. He wants to reduce dependency on imported medicine, for example, by supporting scientists who develop generic drugs.

But Del Bono insists that the emphasis on applied research won't be allowed to hurt basic research at CONICET, which he calls "the spinal cord of our system".

Scientists are cautiously optimistic about the new team. "We are all hopeful," says Conrado Varroto, director of Argentina's space agency in Buenos Aires. "The new authorities, in principle, recognize the importance that science and technology has for the development of the country." ■

♦ [www.conicet.gov.ar](http://www.conicet.gov.ar)

♦ [www.secyt.gov.ar](http://www.secyt.gov.ar)



# Monsoon rains start to ease India's drought

K. S. Jayaraman, New Delhi

A heatwave that scorched most of India throughout May, killing more than 1,400 people, is finally abating with the arrival of the first monsoon rains last week.

In the southern state of Andhra Pradesh, where temperatures reached a record 51 °C late last month, the first heavy rain was expected this week. It was desperately needed, as all nine of the state's main reservoirs had either dried up or reached precariously low levels.

The state government has set up an expert panel, chaired by C. V. V. Bhadrani, director of the Andhra Pradesh meteorological office in Hyderabad, to identify factors that contributed to the drought and to marshal a response to it.

A few weeks ago, the Indian Meteorological Department (IMD) in New Delhi predicted a "normal" monsoon, using a new statistical model; the previous model led to wrong forecasts being made in 2002 (see *Nature* **418**, 713; 2002). But the monsoon that arrived last week was weaker than normal, and hit the north-east of India, rather than the south coast.

Given the unreliability of the electricity supply in many parts of India, the heatwave has wreaked havoc in hospitals and laboratories, many of which lack emergency generators to power air-conditioners and the refrigerators that store vaccines, reagents and chemicals.

Debi Sarkar, a biochemist working at the University of Delhi, points out that "most reagents used in India are imported



Scorched earth: India's worst drought for 20 years (inset) left people struggling to find water.

and kept in customs warehouses until they are collected". He predicts that many will have been degraded during the heatwave.

Subhash Arya, a vaccine expert at the Centre for Logistical Research and Innovation in New Delhi, says that India's polio-eradication programme has been seriously set back by the weather.

India suffers from summer heatwaves from time to time, but according to several

data sources, this summer's drought is the worst for at least 20 years. The heatwave was fuelled by hot air flowing in from the Iranian desert — India normally receives a flow of cool, moist air from the Arabian Sea during May.

Satellite data (see graph) collected by the US National Oceanic and Atmospheric Administration (NOAA) show that more than two-thirds of India's vegetation is under either "severe" or "extreme" drought conditions, says Felix Kogan, a remote-sensing specialist at NOAA's environmental-data division at Camp Springs, Maryland.

An analysis by Uday Shankar De, a geophysicist and former research chief at the IMD in Pune, suggests that the number of hot, dry days in India during May and June has been increasing steadily over two decades. De and colleagues attribute the trend to global climate change.

"The increase in extreme events, such as the heatwave, is linked to global warming," claims J. Srinivasan, a professor of atmospheric sciences at the Indian Institute of Science in Bangalore.

The monsoon rains may not put an end to the crisis, researchers warn. Last year, the monsoon followed its normal pattern until late June and then faltered, resulting in a resumption of drought conditions. "After what happened last year, nobody will dare to make any prediction about the monsoon," says Srinivasan.

De takes a more sanguine approach. "The monsoon is like a train," he says. "If it arrives late at one station it doesn't matter — it will make up for it before it arrives at the next one."

Additional reporting by Hannah Hoag in Washington.

## Quake triggers research expedition

Declan Butler, Paris

Last month's earthquake in Algeria has prompted European researchers to mount a rare full-scale expedition to a region where political conflict over the past decade has sharply curtailed their fieldwork.

The earthquake, which struck the coast on 21 May, killing more than 2,000 people, prompted a French team of 30 Earth scientists to travel to Algeria to collaborate with colleagues at the Centre for Research in Astronomy, Astrophysics and Geophysics in Algiers. At 6.7 on the Richter scale, the earthquake was the worst in the region for more than 20 years.

The North African country lies on the southern side of the fault line at which the Eurasian and African tectonic plates meet, making it interesting to seismologists.

The researchers hope to obtain valuable data about that side of the fault, which has been neglected for more than a decade, says Alain Mauffret, a marine seismologist at the University of Pierre and Marie Curie in

Paris. Because of the conflict in Algeria, many Earth scientists have emigrated from the country, and research by outsiders has been difficult. "We can't send people there to have their throats cut," he says.

Mauffret, who helped to coordinate the mission for France's National Centre for Scientific Research (CNRS), says the need to collect data on the earthquake and its aftershocks made it vital to visit Algeria. The team has installed seismological monitoring equipment, and has measured more than 1,000 aftershocks since 21 May. Underwater stations are due to be added this week.

Satellite radar images are being used to measure ground deformation, and Global Positioning System stations are being deployed to detect ground movements. A hypothesis that the earthquake was triggered by an underwater landslide will be studied by a marine geophysics research group in August, organized by the CNRS and IFREMER, France's national marine research agency.

## Physicists doubt that 'corking' could help baseball's big hitter

**Geoff Brumfiel, Washington**

Baseball fans are starting to question Chicago Cubs superstar Sammy Sosa's impressive hitting record, in the light of his use of an illegal 'corked' bat to boost his slugging power. But physicists say that the bat did little, if anything, to improve his hitting.

Sosa, who has hit 505 home runs, was caught in the act when his bat shattered in a game on 3 June against the Tampa Bay Devil Rays. He denies intentionally using the bat in the game — he claims he uses it to impress fans during practice — but was suspended for eight games.

Physicists say, however, that Sosa gained almost no advantage by using the bat, which was hollowed and filled with a small amount of cork. Using a lighter bat would help Sosa swing harder but would actually lessen the energy transferred to the ball, according to Alan Nathan of the University of Illinois at Urbana-Champaign, who recently studied ball-bat impacts (*Am. J. Phys.* 71, 134–144; 2003). "If there is a net advantage, it would be small enough that it wouldn't be worth writing home about," Nathan says.

Robert Adair, a professor emeritus at Yale University and former official physicist of Major League Baseball, agrees that the perceived advantage of corking a bat is largely "superstition". It is deemed illegal because traditionally bats are supposed to be solid wood. "Corked bats aren't a mortal sin," he says. "Maybe just a venial one." ■

## Prairie-dog model offers hope of tackling monkeypox virus

**Jonathan Knight, San Francisco**

A leading poxvirus expert says he envisages a new line of attack for biologists studying the monkeypox virus in the current outbreak in the midwestern United States.

Mark Buller of Saint Louis University in Missouri says he will try to develop a laboratory model for monkeypox infection using prairie dogs, common rodents that seem to be capable of transmitting the virus.

Until this month, monkeypox — a milder cousin of smallpox — had been seen only in parts of west Africa, where it is endemic in wild rodent populations that are sometimes hunted for food. Now 33 people in Wisconsin, Illinois and Indiana have contracted the disease from pet prairie dogs. The virus is thought to have arrived in a Gambian giant rat, which was imported as part of the growing trade in exotic pets.

"We know nothing about the natural biology of monkeypox at all," Buller says. Research has been limited in part because the disease is rare in humans, but also because there is no good animal model.

Among the animals infected experimentally with monkeypox is the cotton rat, a shoe-sized rodent found in Mexico and the southern United States. These animals succumb quickly to an injection of virus and are useful for testing antiviral drugs and for studying the virus's lethal effects.

But Buller says that a natural host would be much better for studying basic questions about monkeypox, such as how it is transmitted, what cells it infects, what genes control its virulence, and how the host responds. ■



Prairie dogs can transmit monkeypox and are already used to study some infectious diseases.

"You could get a natural profile of the disease," he says.

Buller hasn't ruled out trying the Gambian giant rat, but many animals fare poorly in captivity, and setting up a new laboratory colony of an untried rodent species is fraught with complications. "Some animals turn out to be more trouble than they are worth," he says. Prairie dogs, on the other hand, handle captivity quite well, and are already being used in a number of laboratories studying infectious disease.

Other poxvirus researchers are eager to see Buller's efforts succeed. Donald Smee, a virologist at Utah State University in Logan, would like to test his hypothesis that drug-resistant monkeypox viruses are also less deadly, but has not yet found an animal model in which to do this. Prairie dogs may work if the virus proves lethal to them in the first place, he says. ■

## Pumping row erodes hopes for underground lab

**Geoff Brumfiel, Washington**

Plans to build an underground physics laboratory in an abandoned gold-mine in South Dakota are close to collapse this week amidst quarrelling between its advocates, the National Science Foundation (NSF), and its owner, mining company Barrick Gold.

Barrick says it will now allow the Homestake mine near Lead to begin filling with water, claiming that it was shut out of an NSF panel charged with determining the best conditions for a deep underground laboratory (see *Nature* 423, 578; 2003). A meeting on 9 June between company officials and South Dakota's congressional delegation failed to persuade the company to reverse its decision.

Congressional sources say that Barrick was displeased with the NSF panel report,

released last month, that endorsed Homestake as the best site for the lab. The report warned that if Homestake filled with water, which flows naturally into the mine at a rate of several hundred gallons per minute, this could substantially raise the cost. The panel said it was "unanimous in the opinion that continuing to pump is the most desirable option".

A spokesman for Barrick, which has been under a court order to spend about \$250,000 a month to pump out the mine, denied that the company was angry at the recommendation. But the spokesman, Vince Borg, did say that the company was not allowed to present its case to the NSF panel. "We had up-to-date, accurate, current information we wanted to share with them," he says. "We

didn't understand the decision of the panel."

Within three days of the report's release on 30 May, Barrick announced that it would let the mine flood. This move triggered a letter from many leading physicists, including Stephen Hawking of Cambridge University, urging Barrick to keep pumping.

Some in Congress see the mine as a boon to Barrick and a blight to the federal government. Under a 2001 law, the government must assume all environmental liability for the mine if a laboratory is built there.

But Alfred Mann, a physicist emeritus at the University of Pennsylvania in Philadelphia, is still hopeful that an agreement can be worked out. "This lab would be one of the most important underground laboratories in the world," he says. ■



# Car-safety lobby on collision course with researchers

Tony Reichhardt, Washington

After years of fighting off interference from radio transmitters and broadcast satellites, astronomers are worried that an emerging technology — collision-avoidance radar for cars — could pose the biggest threat yet to some of their observations.

The astronomers' concerns are this time shared by Earth scientists, who say that the radars could mess up their observation of the atmosphere from satellites. But they face an uphill battle in confronting the commercial interests that support the car radars — not to mention the universally popular goal of increased road safety.

So far, only a handful of luxury cars, including models produced by Cadillac, Mercedes-Benz and BMW, offer onboard radar systems that alert drivers when other cars come too close. But their use is expected to widen in the next decade, and US and European regulators are already planning to allocate a share of the radio spectrum to it.

A ruling by the US Federal Communications Commission (FCC) last year allowed the use of car radars at a frequency of 24 gigahertz (GHz). But both NASA and the National Oceanic and Atmospheric Administration opposed that frequency allocation on the grounds that it could interfere with scientific observations.

For radio astronomers, the 24-GHz band is an important frequency for detecting ammonia — a common extraterrestrial molecule. Studies show that millions of car radars operating at that frequency would “absolutely guarantee to close down the whole band” for astronomy, says Wim van Driel, a radio astronomer at the Paris Observatory.

For climate researchers, the frequency sits squarely in a band that serves as a signature of water vapour in the atmosphere.

John Zuzek, who works on spectrum management issues at NASA's Glenn Research Center in Ohio, says that although other agencies were consulted about the FCC rule, studies showing interference problems with scientific observations were essentially ignored. Zuzek and other researchers say that the FCC is under strong pressure from industry to approve the commercial allocations. “We were kind of forced to go along with what happened, but with a lot of reservations,” he says.

Eventually, car radar systems operating at higher frequencies — around 77 GHz — are expected to emerge, but these are currently more expensive to make. European regulators are therefore recommending a phased approach, first using the 24-GHz systems, then phasing in the 77-GHz



systems as the devices become more common in vehicles.

Shifting to the higher frequencies would suit astronomers — although echoes of 77-GHz signals could still be picked up by sensitive radio dishes listening at the 226–231.5-GHz band, one of the ‘windows’ in the atmosphere that is transparent to radio observers on the ground.

Several other issues of concern to researchers will be on the agenda at the World Radiocommunication Conference (WRC-03), a giant round of international talks on dividing up the radio spectrum, which kicks off in Geneva this week. Among the most pressing items are allocations in the 5-GHz range for wireless computer networks. If these proliferate, they could interfere with satellite-based imaging radars used for Earth observation.

Historically, astronomers have been fairly successful at winning protection from encroachment by commercial radio services. And scientists took heart last November when an FCC spectrum task force reaffirmed that “radio astronomy may need to have dedicated, protected spectrum bands for the foreseeable future, due to its highly sensitive applications and the fact that its benefits accrue to society as a whole”.

Yet the combination of scarce spectrum, new uses of the air waves, and the anti-regulatory slant of the current FCC makes this ever more difficult. Astronomers like their mobile phones and wireless Internet connections as much as anyone else, says van Driel, who chairs the European Science Foundation's Committee on Radio Astronomy Frequencies. And he admits that appearing to oppose a technology that could improve road safety “makes us look bad”. But in the scramble for spectrum, he insists, “we have the right to be protected”.

## Pollution fears put work at neutrino laboratory on hold

Nicola Nosengo

Italy's nuclear physics research authorities have temporarily shut down an underground laboratory in a bid to resolve environmental questions that surround its work.

Almost all research at the Gran Sasso National Laboratory, an underground neutrino lab in central Italy, was halted by the Italian National Institute for Research in Nuclear and Subnuclear Physics (INFN) on 29 May.

The move follows warnings from local officials that toxic chemicals used in the lab could pollute local water supplies. Enzo Iarocci, the INFN president, says the closure is intended to draw attention to the problem, which he believes the national government must solve.

The lab has attracted protests since experiments began a decade ago. Environmental groups argue that the toxic chemicals used in the neutrino detectors could leak into aquifers below the lab. Their protests gathered pace last August, when around 50 litres of trimethylbenzene, a chemical that is toxic in high doses, were spilt during work on the Borexino experiment, which studies low-energy solar neutrinos.

The chemical entered the lab's drainage system, and local environmental officials subsequently detected it in a nearby river, although it was not found in drinking water. The local public prosecutor opened an inquiry into safety procedures at the lab, and the regional government set up a parallel expert inquiry staffed by water engineers, public-health officials and INFN representatives.

After assessing contamination risks, the prosecutor ruled on 29 May that the Borexino experiment area should be sealed up. Iarocci, having been told earlier that week by the expert group that the lab's drainage system could be leaking, the same day suspended all experiments involving liquid materials.

Iarocci says the repairs needed are the national government's responsibility. “The INFN is also trying to give a sign of cooperation to local communities,” adds Carlo Gustavino, a Gran Sasso physicist.

Alessandro Bettini, the lab's director, says the national government has reacted positively to the INFN's action, and that work at the lab could restart in around four months, as the repairs needed to the drainage system are minor.

www.lngs.infn.it

## Cities' response to attack hampered by unknown airflow

**Washington** Lack of knowledge about how skyscrapers, parks and urban sprawl affect city airflow could hamper the emergency services' response to a chemical, biological or nuclear attack, the US National Academy of Sciences said last week.

In a report released on 2 June, the academy suggests that meteorological modellers need to do further work on the flow of hazardous agents around cities. The authors say that the models need to provide emergency teams with more information on the probabilities associated with specific predictions.

The report calls for more urban field research and wind-tunnel simulations, better use of meteorological monitoring networks, and unmanned airborne vehicles to supplement fixed observation equipment.

♦ [www.nas.edu](http://www.nas.edu)

## Massive cathedral testifies to ancient emperor's might

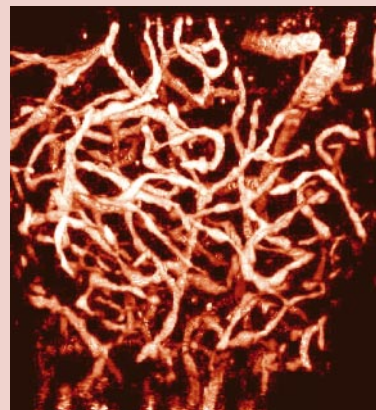
**Munich** German archaeologists believe they have discovered the legendary cathedral of the Holy Roman Emperor Otto the Great (AD 912–973), one of the largest churches

## Laser to shed light on delicate tissue

**Washington** A laser imaging technique could give researchers easy access to three-dimensional images of brains and other organs, its creators say.

Jeff Squier of Colorado School of Mines in Golden used the process to produce this image of blood vessels in a mouse brain. He used short laser pulses to excite fluorescent dyes attached to the vessels, and digitally recorded the light emitted. He then used a second set of laser pulses to burn the upper layer of tissue away, and repeated the process until the whole sample had been scanned.

Squier, who presented the results at last week's Conference on Lasers and Electro-Optics in Baltimore, Maryland, says that the process should be suitable for imaging delicate tissue samples such as embryos.



built north of the Alps during the Dark Ages.

Otto, who reigned over most of central and southern Europe, built the cathedral in Magdeburg, now in eastern Germany, as a show of power. The cathedral's exact site had been a mystery, although it was thought to lie beneath the city's existing gothic cathedral.

The researchers found the ruins, including foundations and graves, when searching a site, about 40 metres from the present cathedral, which some archaeologists believe is the location of Otto's palace. "The graves

we found at the north and south foundations are typical for churches and it definitely rules out the palace theory," says Rainer Kuhn of the Archaeology Heritage Service in Saxony-Anhalt, who led the dig.

## SARS anxiety disrupts conferences

**Washington** Scientists from areas affected by severe acute respiratory syndrome (SARS) are to be screened by telephone interview



before attending a major cancer conference in Washington next month.

Organizers of the annual meeting of the American Association for Cancer Research (AACR) say that people from countries for which the World Health Organization has issued a travel warning will be asked whether they have had contact with SARS patients or suffered SARS-like symptoms. Those answering yes will need a doctor's clearance to attend. The meeting was to have been held in Toronto in April, but was rescheduled three days before it was due to open because of a SARS outbreak in the city (see *Nature* 422, 547; 2003).

A new spate of SARS cases in Toronto led organizers of a separate conference — the International Union of Biochemistry and Molecular Biology's 19th annual congress, which was scheduled for July — to cancel their meeting. Organizers say a successful meeting could no longer be held because too many registrants had already dropped out.

## Charities chip in to save chromosome institute

**Washington** The Eleanor Roosevelt Institute, a key player in the sequencing of human chromosome 21, has been saved from closure thanks to help from two private foundations and the University of Denver.

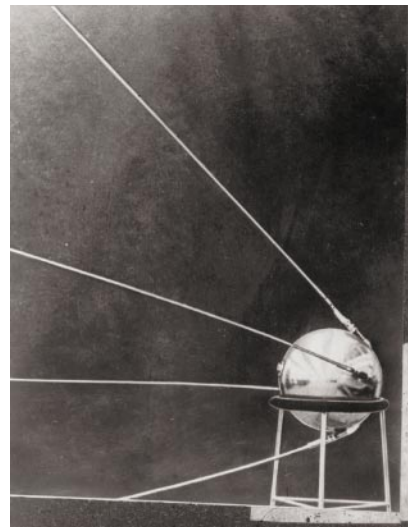
**The non-profit institute, which is based in Denver, Colorado, looked doomed in April after a downturn in charitable donations and the loss of a large grant from the National Institutes of Health (see *Nature* 422, 462; 2003). But two Denver-based charities — the Bonfils-Stanton Foundation and the Boettcher Foundation — have chipped in \$970,000 to pay off the foundation's building mortgage, and the institute has now merged with the city's university.**

## First satellite for sale, one careful owner

**Washington** Do you want to own a piece of space history? This could be your chance. A US classic-car dealer is selling what he claims to be an original version of the Soviet Sputnik 1 satellite.

The satellite is a souvenir from the dawn of the space age, says George Stauffer of Stauffer Classics in Blue Mounds, Wisconsin, who is asking for US\$39,000 for the item on his website. Sputnik 1, the first man-made satellite to orbit the Earth, was launched in October 1957. Less than 60 cm high, it caused gnashing of teeth among US politicians, who feared that they were falling behind their Soviet rivals.

Interest in Sputnik surged after the online auction company eBay advertised a similar



**Space pace-setter: Sputnik, shown resting on a three-legged pedestal for its first official photo.**

sale last week, although the item has since been removed following a flurry of fraudulent bids. Historians say that Soviet scientists built between 4 and 20 more models, for use in testing, demonstrations and as diplomatic gifts. Nobody can say for sure how many authentic Sputniks exist, although many museums claim to own one.

► [www.staufferclassics.com](http://www.staufferclassics.com)

AP/TASS



The former Soviet Union's network for bioweapons research crossed the country (left), but international efforts are now trying to redeploy these facilities for peaceful means.

V. ZININ/ITAR-TASS

## Still out in the cold

Collaborations between Western researchers and former Soviet bioweapons scientists could benefit both parties. But mistrust and bureaucracy are getting in the way, says Geoff Brumfiel.

In autumn 2001, three American researchers sped down a deserted two-lane road that cuts through the forests south of Moscow. They were travelling to Obolensk, once a secret city and home to one of the former Soviet Union's largest bioweapons research complexes — the State Research Center for Applied Microbiology.

The researchers were part of a programme, funded by the Pentagon, that aims to keep Russia's former bioweapons scientists gainfully employed on useful projects. Despite the dilapidated surroundings in Obolensk, the visitors were enthused by the opportunities for collaboration. The crumbling concrete buildings “looked almost like a ghetto”, recalls Rebecca Morton, a veterinary scientist at Oklahoma State University in Stillwater. But after two weeks, she had hatched a plan to work with Obolensk researcher Vitaly Pavlov on endemic Eurasian strains of *Francisella tularensis*. This bacterium causes tularaemia, a potentially fatal and extremely infectious disease that affects the liver, spleen, lungs and lymph nodes, which was studied at Obolensk because of its bioweapons potential.

Morton's project, which would study the surface proteins on different strains of the bacterium in an effort to develop strain-specific diagnostic tests, is exactly the sort of initiative that the programme is designed to support. But almost 18 months down the

line, she is no nearer to getting the project under way. Although her proposal has had positive peer review, the funding request is still winding its way through the Pentagon's bureaucracy. “I haven't spoken to Vitaly for a while, because I don't have much to tell him,” says Morton. “It's a little embarrassing.”

### Obstacle course

Other researchers who hope to set up collaborations at the Obolensk centre and its sister facility, the State Research Center of Virology and Biotechnology, known as Vector, at Koltsovo near Novosibirsk in Siberia, are experiencing similar delays. Cultural differences, mistrust between Russia and the United States, and bureaucratic obstacles on both sides are all conspiring to stall promising avenues of research.

After the demise of the Soviet Union in 1991, US funds flowed rapidly to former nuclear scientists and rocket engineers, with the goal of preventing them from accepting lucrative offers from countries eager to acquire an arsenal of ballistic nuclear weapons. But bioweaponers were left out in the cold. The reason, according to Amy Smithson, a senior associate at the Henry L. Stimson Center — a security-policy think-tank in Washington DC — was that US officials lacked contacts inside the super-secret Soviet bioweapons network. “The biological

non-proliferation programme literally had to be started from scratch,” she says. As a result, more than half of the staff at Obolensk and Vector melted away during the 1990s — to where, no one knows for sure.

Obolensk and Vector were two research powerhouses in a network of facilities spread throughout the Soviet Union, known collectively as ‘Biopreparat’ (see map, above). This network weaponized diseases such as plague, anthrax, tularaemia, brucellosis and smallpox, behind the façade of a state-run pharmaceutical enterprise. Scientists at Obolensk and Vector even genetically engineered bacteria to resist antibiotics. In addition to the staff's expertise, the centres have containment labs for working on dangerous pathogens — the provision of which is currently a limiting factor in US plans to ramp up biodefence research.

With Russia now suffering epidemics of diseases such as tuberculosis and AIDS, it stands to benefit from projects that would redirect the expertise at Obolensk and Vector to these problems. “Russia is a time bomb right now,” says Ann Harrington, who studies options for reducing the threat of bioweapons at the National Defense University in Washington DC. “It has an enormous need for facilities that can support public health, and that can monitor and identify disease.”

In her former job as acting director of the

## Blazing the trail

V. ZININ/TAR-TASS

Jens Kuhn describes himself as a “microbe nerd”. Having a particular fascination with the Ebola virus and related killers — known as haemorrhagic viruses — this young German virologist was delighted to become the first Western scientist to work at the bench in the heart of the former Soviet bioweapons establishment.

Kuhn’s “good fortune”, as he puts it, began in autumn 2000 when he found himself chatting with an official from the Pentagon at the inauguration in Stockholm of Sweden’s first level-4 containment facility for handling dangerous pathogens. In his short career, Kuhn had already worked in South Africa, South Korea and at the US Army Medical Research Institute of Infectious Diseases in Fort Detrick, Maryland, where he had spent a year, until April 1999, preparing a master’s thesis on the development of Ebola vaccines.

Kuhn had also started to survey the open literature on Ebola, and had contacted researchers at Vector, the State Research Center of Virology and Biotechnology near Novosibirsk in Siberia, who had begun publishing, in Russian, in 1992. He was surprised to get a swift answer from this formerly secretive bioweapons lab. “They were just as keen for acknowledgement of their work as any other scientists,” says Kuhn. The contact spawned the idea of a visit to Vector, “which seemed impossible at the time”, he says.

### Lucky break

But thanks to his meeting with the Pentagon official, that mission suddenly became possible. The official recognized that Kuhn — young, intelligent, idealistic yet experienced — was exactly the sort of person to serve as a test case for a new Pentagon-sponsored programme, administered by the International Science and Technology Center (ISTC) in Moscow, to send young scientists to work for extended periods in former Soviet bioweapons labs, to help support the labs’ conversion to peaceful projects.

Within nine months of that chance meeting, Kuhn was at Vector as a temporary employee of the Science Applications International Corporation, a contractor for the US Department of Defense. Vector’s buildings and equipment were ramshackle, Kuhn says, but his welcome from the lab’s rank and file was warm. “The scientists supported me in what I wanted to do, and also invited me to all their parties and celebrations,” he says. “They treated me like a colleague although I was only a student.”

But Vector’s leaders seemed more suspicious. Kuhn spent his first five weeks in Siberia locked out of the lab, continuing a debate about which research project he should undertake, and overcoming numerous bureaucratic obstacles. He had medical checks with 12 doctors, including a psychiatrist who spent three hours determining if he was mentally stable. He had to study for, and pass, a biosafety exam in Russian. “I don’t suspect it to be delaying tactics but we’ll only



**Building bridges:** Jens Kuhn (inset) was the first Western scientist to study at the State Research Center of Virology and Biotechnology in Siberia.

know for sure when the next Westerner arrives,” comments Kuhn, who for the benefit of future visitors has prepared a list of the medical checks that can be done before arrival and an English translation of Vector’s biosafety protocols.

Once in the lab, Kuhn made rapid progress with his project despite some eccentric working conditions. Contrary to fears he had heard expressed in the West, the pathogen containment itself was secure. “The Russians don’t want to kill themselves any more than Western scientists,” Kuhn says. Any safety problems arose from smaller economies, such as the absence of a red light in the darkroom, which meant stumbling blindly while holding biological samples, electrophoresis equipment that threatened electric shocks to the unwary, and the failure to repair uneven stairs. “I dropped one of my immunoassays the first time I went downstairs to read my samples,” says Kuhn.

### Screen star

Thanks to Kuhn’s reports back to the ISTC, conditions improved over time. More rubber gloves and pipette tips were shipped to Vector, for instance, after Kuhn found his Russian colleagues removing his discarded supplies from the bin.

Over the course of some four months, Kuhn studied the Crimean-Congo haemorrhagic fever virus, which causes hundreds of infections each year in central Asia. He screened some 1,500 samples taken from patients, autopsies and ticks, the carriers of the disease, which had been collected from all over the former Soviet Union,



and characterized seven distinct viral strains. “The project was not ‘fancy’,” says Kuhn, “but was absolutely essential for understanding the epidemiology of strains, which is necessary for developing vaccines and diagnostic kits.”

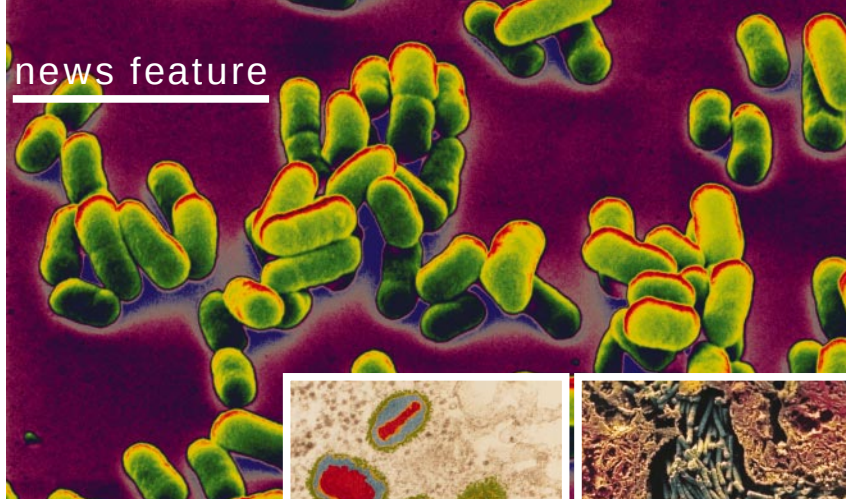
Kuhn, who is now looking for a postdoctoral position, believes that other virologists have much to gain from following his lead. Despite a tendency towards descriptive research, with numerous parameters but few solid conclusions, Kuhn is convinced that Vector’s scientists have what it takes to compete internationally. “All they need is the right equipment,” he says.

“My time at Vector was well-spent,” says Kuhn, “both in terms of life experience and in terms of science.” Indeed, he hopes to make a more extended visit in future, should the ISTC-administered programme provide further funds. “It’s not only my passion for the haemorrhagic bugs they have there,” Kuhn explains. “I also like the idea of helping to make this world a little bit safer by moving ahead the transformation of Vector to a civilian centre.”

Alison Abbott



## news feature



Russian scientists have expertise to offer the West in studying pathogens such as plague bacteria, smallpox viruses (right) and anthrax bacteria (far right), which were all weaponized by the Soviet Union.

US state department's Office of Proliferation Threat Reduction, Harrington helped to set up the programme that took Morton and her colleagues to Obolensk. The *modus operandi* of this scheme, funded by the US Department of Defense and administered by the US National Academies, is to build partnerships between Western academics and the former Soviet bioweapons establishment. A sister programme, run by the International Science and Technology Center (ISTC), a non-proliferation organization in Moscow funded in most part by the European Union, Japan and the United States, aims to pay for more extended visits by foreign scientists to former Soviet bioweapons labs (see 'Blazing the trail', previous page).

### Take your partners

Under the National Academies scheme, researchers are paid to travel to Obolensk or Vector for up to two weeks in search of partnerships. If they find a Russian to work with, they draw up a proposal, which is reviewed by the US academies' National Research Council, and can win up to \$10,000. The idea is that participants will then apply for further grants from the Pentagon, the ISTC or other non-proliferation bodies.

Researchers selected for the first round of visits went to Russia in autumn 2001. "I had this idea that I could essentially extend my lab and also switch to interesting organisms that I wouldn't be able to study in the United States," says Konstantin Severinov, a microbiologist of Russian extraction who now works at Rutgers University in New Jersey.

But so far, little progress has been made towards realizing the programme's potential. At a National Academies meeting in Washington DC last December, Severinov and Gregory Ebel, an immunologist with the New York State Department of Health in Slingerlands, expressed their frustration.

Severinov, who studies viruses called phages that infect bacteria, said that his research at Obolensk has slowed to a crawl, and Ebel explained that both US and Russian customs officials were blocking transport of even the most simple equipment.

The problems have several causes, but many stem from the secretive culture of Biopreparat. For decades, the network's scientists were cut off from the outside world, other Russian researchers and even each other. Unsurprisingly, they are not familiar with the grant writing, publication and peer review that underpins mainstream science.

At higher levels, trust continues to be an issue. Many senior managers at Obolensk and Vector are veterans from the Soviet era and have a deep mistrust of the United States. They have almost absolute authority over their labs—determining what can flow to the West, and having an obligation to the Russian state to protect classified research. "Scientists may be convicted for giving state secrets to foreigners," says Ken Alibek, who served as deputy director of Biopreparat for five years before defecting to the United States in 1992.

As a result of these attitudes, some US politicians complain that the Russian labs are simply trying to take cash handouts without opening themselves up to proper scrutiny. "We must ensure that the investment can be directly traced to an actual tangible reduction in military threats," the chair of the House of Representatives Armed Services Committee, Duncan Hunter (Republican, California), said at a hearing in January. Suspicions about the new schemes are heightened by the experience of earlier non-proliferation programmes established in the former Soviet Union, which have been plagued by corruption: lab administrators have been known to take a cut from each research grant at their facility.

Today, financial checks are in place to prevent such abuses, but these also slow research.

Following a congressional crackdown, for instance, there are now strict limits on how much of the funding can be spent by US researchers on projects in Russia. "We have some money for travel," says Bruce Scharf, a veterinary scientist at the State University of New York's Downstate Medical Center in Brooklyn, who is setting up a project to study rabies at Vector under the National Academies programme. "But I'm not paid a cent."

### Closed borders

Perhaps the most serious problems are those caused by customs and immigration restrictions. Especially since the terrorist outrages of 11 September 2001, and the anthrax attacks that followed, the US customs service has enforced strict controls on the import of biological material. As a result, Scharf has been unable to get the rabies samples he is studying at Vector into the United States. Similarly tough regulations on both the US and Russian sides are preventing Sergey Morzunov, a Russian-born microbiologist at the University of Nevada at Reno, from sending even basic materials, such as reagents for a DNA sequencing kit, to Vector. "The expiry date for my sequencing kit is May 2003, but it is still sitting on a shelf in the warehouse," Morzunov complains.

New immigration regulations have also stopped Russian partners in the programme from visiting the United States to build links with their new Western colleagues, adds Vladimir Volkov, deputy director of the Obolensk facility. "Getting a visa for a business trip to the States may now take over three months," he says.

Given the litany of problems, Alibek doubts whether the programmes will do much to further the cause of non-proliferation—especially as so many Obolensk and Vector staff drifted away in the 1990s. But other experts point to the vast expertise on biological warfare still present at the centres, and argue that it must be worth harnessing this knowledge. "I think the programme is still very much in its infancy," says Glenn Schweitzer, its coordinator at the US National Academies.

Despite the difficulties he has experienced, Severinov remains committed to the goal of drawing the former outcasts of Biopreparat into the international scientific community. He and his collaborators have nearly finished two papers, the first of which he hopes will be published later this summer. And Severinov is now applying for a larger grant that would permit him to extend his collaboration at Obolensk for three more years. "These people must have joined those institutes because they like science," he says. "I think they just have to be put into the mainstream." ■

Geoff Brumfiel is *Nature's* Washington physical-sciences correspondent.

♦ [www7.nationalacademies.org/dsc/biomedical\\_exchange\\_program.html](http://www7.nationalacademies.org/dsc/biomedical_exchange_program.html)



# Gas leak!

Global warming isn't a new phenomenon — sea-bed emissions of methane caused temperatures to soar in our geological past. But no one is sure what triggered the release. Quirin Schiermeier investigates.



D. HULSHIZER/AP

About 55 million years ago, our planet emitted a spectacular burp. Trillions of tonnes of methane, until then safely locked up in soils and beneath the ocean floor, were released into the oceans and atmosphere. Methane is a greenhouse gas, so the result was a global-warming incident that has not been matched since — within a few thousand years, average temperatures in some areas rose by up to 8 °C.

Although the evidence for this warming is clear, what sparked the methane release is a mystery. Some researchers believe that the trigger was a small change in ocean temperature. Others suggest that a comet impact was responsible. Analyses of cores from ocean-drilling experiments may soon provide the answer, but the results might do more than shed light on ancient climate — they may also tell us what the long-term consequences of current global warming are likely to be.

The ancient warming, known as the Palaeocene/Eocene thermal maximum (PETM), is the most prominent climate episode of the past 65 million years. It was first identified in the late 1980s by oceanographers analysing oxygen isotope levels in a core drilled from the sea bed near Antarctica<sup>1</sup>. Water molecules containing the lighter oxygen-16 isotope evaporate more readily than those containing oxygen-18, so the relative proportion of oxygen-18 in the oceans rises if the temperature

increases. Sedimentary rocks containing shells and plankton that formed in warm water also tend to be richer in oxygen-18. In the case of the Antarctic core, the ratio of the two oxygen isotopes in the sedimentary rocks showed that the sea surface temperature was about 12 °C before the PETM. But sometime in a 10,000-year window around the PETM, they leapt by 7–8 °C, and didn't return to pre-PETM levels until about 250,000 years later.

## A sea change

Researchers wondered what could have caused such a jump. Hints lay in carbon isotopes extracted from the same marine sediments used to identify the PETM. Living organisms such as algae take up carbon-12 in preference to carbon-13 during photosynthesis, so the proportion of carbon-12 in sea water rises when less photosynthesis happens. This has a knock-on effect — the shells of marine animals that form in water where

**The mechanisms that helped bring down temperatures in the past are still active today.**

**Digging deep: the ship *JOIDES Resolution* has collected sea-bed samples that may shed light on the climate conditions endured by phytoplankton (inset) millions of years ago.**

little photosynthesis is taking place also contain relatively high levels of carbon-12.

Sedimentary rocks containing shells dating from the PETM feature high amounts of carbon-12, but the levels are so high they cannot be explained even if all photosynthetic activity had ceased when they were formed. Researchers concluded that there must have been a massive release of carbon-12 into the ocean and atmosphere — an injection similar to, or perhaps exceeding, the total amount of carbon dioxide and other greenhouse-gas releases pumped into the atmosphere since the industrial revolution.

An explanation for the carbon-12 spike, now known as the methane-burp theory, was suggested in 1995 by Gerald Dickens, a geologist then at the University of Michigan in Ann Arbor<sup>2</sup>. He argued that the carbon came from methane hydrates — ice-like crystalline solids in which methane molecules are trapped inside frozen water. Then, as now, hydrates were found in Arctic tundra soil and the deep oceans near continental slopes. The gas fits the bill as it causes greenhouse warming, and the methane in hydrates is rich in carbon-12.

If the methane escapes from its icy cage,

A. SYRED/SPL

it can combine with dissolved oxygen in the ocean, in a reaction probably driven by bacteria, to produce carbon dioxide and water. Some of the carbon dioxide will escape to the atmosphere, along with some unreacted methane, where both act as greenhouse gases.

Hydrates are only stable in a narrow range of pressures and temperatures, so any ocean warming could disrupt these conditions, releasing more methane. Just a small amount of warming could kick-start a positive feedback loop between hydrate release and further warming, sending global temperatures soaring. According to recent work, 1,500–2,000 gigatonnes ( $10^9$  tonnes) of carbon, rich in the lighter isotope, would need to have been added to the ocean–atmosphere system in less than 10,000 years to cause the temperature rise and create the distinctive carbon isotope ratio seen for the PETM<sup>3</sup>.

## Bubbling under

Evidence for the methane-burp theory comes from studies of Blake Nose in the subtropical Atlantic off the coast of South Carolina. These indicate that 55 million years ago a major sea-bed landslide occurred in the region<sup>4</sup>. Methane release, which would have destabilized the ridge, is one possible cause. The burp theory is also backed by climate simulations. This January, for example, Gavin Schmidt and Drew Shindell, atmospheric chemists at Columbia University in New York, published a simulation confirming that the amount of methane thought to have been released from hydrates is consistent, in terms of its lifetime in the atmosphere and its greenhouse effects, with the estimated temperature changes at different latitudes during the PETM<sup>5</sup>.

But this still begs the question of what prompted the methane release. A recent suggestion by Dennis Kent, a geologist at Rutgers University in New Jersey, invokes the impact of a comet containing large amounts of carbon-12 (ref. 6). Kent says that the impact would have vaporized parts of the comet, adding carbon to the atmosphere and triggering methane release from the hydrates.

The idea has some intriguing evidence to support it. For example, Kent points out that there is a jump in the levels of iridium found in fossils dating back to the PETM<sup>7</sup>, and that comets are often rich in iridium. But one crucial part of the story is missing — the crater that such an impact would have left, although it could be hidden on the ocean floor.

Other researchers argue that a change in ocean conditions could have caused the release<sup>8</sup>. Global sea surface temperatures were increasing by about 0.5–1.0 °C every million years at the time of the PETM. This rise could have reached a threshold level that altered ocean circulation, channelling destabilizing warm water towards the hydrates.

So which mechanism was the trigger? It is still unclear exactly what “uncorked the



**Caged in:** methane trapped in icy hydrates may have caused global warming millions of years ago.

bottle”, says Miriam Katz, a marine geologist at Rutgers University. Katz has developed models to see whether a change in ocean circulation could release enough methane to start the warming<sup>9</sup>, or whether other causes, such as sea floor landslides, could have triggered an abrupt degassing of methane reservoirs<sup>4</sup>. “With existing data, neither mechanism can be identified unequivocally as triggering methane release,” she says.

## Core issue

But new information should soon be available. Last month, the US drilling ship *JOIDES Resolution* completed a two-month expedition to the Walvis Ridge, off the coast of Namibia. Sediment cores were recovered from five sites, and these should provide a better picture of conditions at the time of the PETM. By using these data to improve their models, researchers hope to rule out all but one of the possible trigger mechanisms.

Apart from the cause of the PETM, climatologists are also interested in its impact, and how the warming was reversed. Studying the PETM, says Ursula Röhl, a marine geologist at the University of Bremen in Germany who sailed on the *JOIDES Resolution* expedition, provides climate scientists with an extreme test of climate theory and models. It also helps them to understand how current global warming might affect life on land and in the oceans, she adds.

Although details of the transfer of carbon between rocks, the atmosphere and the ocean were different 55 million years ago — the total amount of carbon dissolved in the ocean was probably greater than it is now, for example — the underlying mechanisms are not. “The basic biogeochemical reactions and processes, including the biological pump that draws down carbon dioxide from the atmosphere to the deep ocean, should be the same,”

says James Zachos, a palaeoceanographer at the University of California, Santa Cruz, who also sailed on the *JOIDES Resolution*.

This biological pump is probably responsible for the gradual return of temperatures to their pre-PETM levels. The process may have been helped by increased growth of plants and plankton<sup>10</sup>. Higher temperatures are likely to have accelerated the physical and chemical erosion of soils and rocks, for example, raising the amount of iron and phosphate washed into the ocean. These nutrients drive photosynthesis, and cause more carbon to be absorbed by organisms in the upper oceans.

The good news, then, is that the mechanisms that helped to bring down temperatures in the past are still active today. Increased amounts of vegetation, ocean phytoplankton and erosion could all act as a natural brake on today’s global warming. But studies also suggest that we shouldn’t rely on these systems to rein in global warming in the short term. “The climate events at the PETM boundary teach us that the Earth can help itself,” says Richard Corfield, a marine geochemist at the University of Oxford. “But we should always bear in mind that it took at least 100,000 years to recover.” ■

**Quirin Schiermeier is Nature’s German correspondent.**

1. Kennett, J. & Stott, L. *Nature* **353**, 225–229 (1991).
2. Dickens, G. R., O’Neil, J. R., Rea, D. K. & Owen, R. M. *Paleoceanography* **10**, 965–971 (1995).
3. Röhl, U., Bralower, T. J., Norris, R. D. & Wefer, G. *Geology* **28**, 927–930 (2000).
4. Katz, M. E., Pak, D. K., Dickens, G. R. & Miller, K. G. *Science* **286**, 1531–1533 (1999).
5. Schmidt, G. A. & Shindell, D. T. *Paleoceanography* **18**, doi:10.1029/2002PA000757 (2003).
6. Kent, D. et al. *Earth Planet. Sci. Lett.* **211**, 13–26 (2003).
7. Schmitz, B. et al. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **133**, 49–68 (1997).
8. Thomas, D. J., Zachos, J. C., Bralower, T. J., Thomas, E. & Bohaty, S. *Geology* **30**, 1067–1070 (2002).
9. Katz, M. E., Cramer, B. S., Mountain, G. S., Katz, S. & Miller, K. G. *Paleoceanography* **16**, 549–562 (2001).
10. Bains, S., Norris, R. D., Corfield, R. M. & Faul, K. L. *Nature* **407**, 171–174 (2000).



# Why should we help politicians to wage war?

Scientists should not fool themselves that their work for the military has ethical benefits.

Sir — Your Editorial “Hope for best, prepare for worst” (*Nature* **423**, 101; 2003) was characteristically illuminating. Given the depressing situation in which the world has been left following the war on Iraq, it is quite reassuring to learn that a strong competitor to President Bush for the title of ‘world’s scariest’ has surfaced in the person of North Korean President Kim Jong-Il.

The Editorial recommends that we scientists should pay special attention and search our own consciences: “scientists in East Asia should be thinking about how they might be involved”. The good news, apparently, is that US scientists, “since the Manhattan Project”, have become engaged in the international military thinking of their governments. The new danger, you say, is that, in Asia, “scientists are adopting the attitude that war is something for the government to deal with” and adopting the view “don’t let political considerations skew your objectivity; focus on your research; be driven by pure curiosity”.

You state that American scientists, since the Manhattan Project, have behaved ethically (if not always successfully) in trying to influence the use that is made of their work. Could you give some examples? Was the large-scale defoliation in Vietnam a success for biologists or ecologists? Is the use of depleted uranium, now so popular in US munitions (for example, in Serbia, Afghanistan and Iraq), an ethical achievement by physicists? Is the dropping of 2,000-pound bombs in residential neighbourhoods due to “leading chemists”?

It is unfortunate that you present India and Pakistan as positive examples of the involvement of scientists in military development. Both countries parade nuclear weapons in the service of rather unbecoming nationalistic and bigoted politics. No dissent by “leading scientists” has become publicly known. And should one mention Israel?

It would also have been helpful had you explained the advantage of involvement by “leading researchers” in military circles. Is this necessary because of the need for ever more sophisticated weapons? Or is it because non-leading scientists may have less ethical backbone? Indeed, could the threat of participation by “those much less qualified” have led to a somewhat better score?

The main disadvantage you ascribe to Asian scientists’ unwillingness to be involved in military work — since “military preparations require science-

based research and development, from the analysis of an enemy’s capabilities to the design and production of weapons” — may be their inability to provide the sophisticated tools that are necessary to inspect and destroy the weapons of some future enemy, and then to decimate the presumed enemy with laser-guided cluster bombs.

Luckily, the threat posed by the attitudes of Asian scientists is reduced, mainly because — following the wars on Serbia, Afghanistan and Iraq — international institutions such as the United Nations, the United Nations Security Council and the Geneva Conventions are no longer available for destruction. Should the current 30 months of war on Palestinian autonomy be thrown into the deal? Was it perhaps the ‘ivory tower’ attitude of Asian institutions of science and culture which prevented any sign of alarm being shown at the devastation of international order? This silence has been particularly deafening in contrast to the persistent and principled stands that have been taken by all the western religious

institutions, at their highest levels.

There is in fact a strong case for the opposite attitude, for lauding those who carry on research for “the glory of God”, beautifully articulated by Sir Michael Atiyah in his address concluding his term as president of the Royal Society in November 1995. “In this semi-political world ... we [the scientific community] are in danger of losing our way and our identity. The scientific ethos becomes increasingly hard to discern. Scientists are too often thought of as a sinister part of the establishment.”

In a world in which rising public opinion is considered by the *New York Times* to be the second world war, science should be better advised — not least when one peruses such ‘academic’ projects as Massachusetts Institute of Technology’s Institute for Soldier Nanotechnologies (see <http://web.mit.edu/isn/>).

**Daniel Amit**

*Racah Institute of Physics, Hebrew University, Jerusalem, Israel, and Istituto di Fisica, Università di Roma, La Sapienza, Ple Aldo Moro 2, 00185 Rome, Italy*

## DNA discoveries through crystallography

Sir — During the years that followed the discovery of the DNA double-helix structure, whose fiftieth anniversary is currently being celebrated (see [www.nature.com/nature/dna50](http://www.nature.com/nature/dna50)), a wealth of fibre diffraction studies were carried out (reviewed in ref. 1).

However, the atomic details of the double-helical structure could not be confirmed until chemically pure oligonucleotides became available in the 1970s and allowed the study of single crystals.

An important breakthrough was the crystallization of an RNA mini-helix<sup>2</sup>. Oligonucleotide crystallography has provided valuable information and unexpected discoveries, yet there is plenty of scope for more.

The two first DNA model structures that were determined produced more surprise than confirmation. The first of these<sup>3,4</sup> was published on 22 June 1978, obtained from d(ATAT) and showing only parts of a double helix in a rather complex crystal structure. A year later, the second structure<sup>5</sup> uncovered left-handed DNA. It was only in 1980 that the right-handed double helix proposed

in 1953 could be seen in atomic detail<sup>6</sup>.

Since then a wealth of crystallographic data have fully confirmed the double-helical nature of B-form DNA and the Watson–Crick pairing of the bases in this and most other forms of DNA.

In 1979 Z-DNA was discovered, followed by quadruplexes, Holliday junctions, Hoogsteen pairing and others, as detailed in the nucleic acid database (<http://ndbserver.rutgers.edu>). This database has been recently improved, and provides images of any available structure.

Some may think the field is mature, with the structures of more than 600 deoxyribonucleotides known. Yet these data are extremely limited; for example, only two structures have been determined with only AT pairs — most structures are rather CG-rich. Some sequences are easier to study, and so about 85% of crystallized B-form DNA models start with cytosine.

**Juan A. Subirana**

*Department of Chemical Engineering, Universitat Politècnica de Catalunya, Diagonal 647, Barcelona 08028, Spain*

1. Leslie, A. G. W., Arnott, S., Chandrasekaran, R. & Ratliff, R. L. *J. Mol. Biol.* **143**, 49–72 (1980).
2. Rosenberg, J. M. *et al. Nature* **243**, 150–154 (1973).
3. Viswamitra, M. A. *et al. Nature* **273**, 687–688 (1978).
4. Viswamitra, M. A. *et al. Biopolymers* **21**, 513–533 (1982).
5. Wang, A. H. J. *et al. Nature* **282**, 680–686 (1979).
6. Wing, R. M. *et al. Nature* **287**, 755–758 (1980).

# Learning to live with AIDS

New approaches may be needed if AIDS is to be controlled.

## Combating AIDS: Communication Strategies in Action

by Arvind Singhal & Everett M. Rogers  
Sage: 2002. 426 pp. £35, \$67.95 (hbk);  
£14.99, \$29.95 (pbk)

## Living with the AIDS Virus: The Epidemic and the Response in India

edited by Samiran Panda, Anindya Chatterjee & Abu S. Abdul-Quader  
Sage: 2002. 204 pp. £27.50, \$45 (hbk);  
£14.99, \$24.95 (pbk)

Ritu Priya

Efforts to control AIDS in developing countries have largely had an international basis. There is a vast literature on AIDS-control strategies, but the most visible reports have been from international agencies, looking at the magnitude of the epidemic and the 'best practices' for its control. Advocacy material and manuals for training programme managers and AIDS workers abound. But only a small segment consists of rigorous analysis of the interventions undertaken so far.

The approach set out during the 1980s has thus become a set of globally communicated homilies that are universally linked with AIDS-control work. These include valuable principles such as respecting human rights in disease control; providing information to all; positing a public-health problem as more of a development issue than a medical issue; establishing partnerships between different social groups; and focusing on the conditions of the socially marginalized sections of the community.

Translating these principles into reality, however, requires consideration of the specific context within which they are to be applied. For instance, the cost of antiviral drug regimens for HIV and AIDS is important in regions where medical care is assured, but elsewhere it can distract attention from the essential issues of delivering basic care and treating opportunistic infections. Likewise, the promotion of sexual freedom may be liberating in some contexts but is detrimental to the rights of women under conditions where they have no other freedoms and so are most vulnerable to exploitation.

National AIDS-control programmes the world over basically follow a similar set of strategies. The programmes need to be based on a contextualized assessment of the existing resources — such as materials, infrastructure, support systems and value frameworks — that are available in each community and in society at large.

*Combating AIDS* draws upon material on AIDS control in South Africa, Kenya, Brazil,



**Local knowledge: considering cultural factors can make communication about AIDS more effective.**

Thailand, Cambodia and India to extract universal lessons for developing countries. Its focus is on communication strategy, but half of the book is devoted to discussing the global history of the epidemic, including its biomedical, clinical, epidemiological and policy dimensions. In these chapters the book is disappointing. The strong chapters are those on cultural strategies and entertainment-education, which draw attention to the limitations of approaches that ignore cultural strengths and the opportunities that they provide for positive intervention. The book's illustrations of innovations with existing non-verbal communication in collectivistic cultures — for example, through dress codes that indicate pregnancy and hence the husband's responsibility to his wife at such times — demonstrate sensitivity to cultures other than those of Western Europe and the United States.

However, the authors' evaluation of communication as an effective intervention for controlling AIDS is not convincing. The evidence is usually anecdotal, and where a field trial is cited, the data for comparison across the study and control groups are not provided. Indices such as self-reported shifts to safer practices are not reliable, as respondents are likely to give the 'expected answers' in the post-intervention period.

In areas of high prevalence, spontaneous

declines in unsafe behaviours as a result of direct exposure to disease and death on a large scale cannot be discounted. In areas of low HIV prevalence, declines in the rates of infection with sexually transmitted diseases do not necessarily indicate that the sustained low HIV seropositivity is due to the communication intervention; other epidemiological factors could be involved. This is not to say that communication to the public is of no value, but that its limitations when the wider environment is not conducive or is even contrary to such intervention need to be recognized.

The book contains many paradoxes. The authors espouse the view that "the forest is more important than the individual tree", yet their own focus remains at the 'micro' level. They critique the emphasis on biomedical dimensions but remain preoccupied with the use of condoms and anti-retroviral drugs as the final goal of AIDS communication. And their blurring of the distinction between the use of spirituality and social values, and the use of the power of religious leaders for promoting responsible sexuality will worry those concerned about human rights.

*Living with the AIDS Virus* presents a more detailed analysis of the effort to control AIDS in India, with contributions from those who have been involved with the campaign over several years. A contextual sensitivity

REUTERS



and an attempt to identify the lacunae in existing approaches are revealed in most of the chapters, especially those on the government's response, the role of non-governmental organizations, community responses in Mumbai City, and interventions with gay men. These chapters demonstrate the complexity of the issues involved, and contrast with the superficial analysis presented by *Combating AIDS*. The negative impact of the information, education and communication strategy in terms of increasing fear and stigma is recognized. *Living with the AIDS Virus* also highlights the limitations of 'targeted interventions' as the target groups tend to be stigmatized. Instead, it is better to address communities at large; for example, rather than target truckers, it is better to reach societies along transport corridors.

Both books acknowledge that AIDS-control strategies have generated stigma, ignored the creation of capacities for care and support, and been gender insensitive. But despite this, both books propose managerial and technocratic solutions that do not question the existing perspective. If we do not want to prescribe without a diagnosis, we need to ask why such measures were adopted in the first place.

Alternative perspectives with a societal rather than a programme-oriented approach have also been articulated since the 1980s. They integrate the biomedical and social dimensions rather than dichotomizing them, and view preventive and curative measures as an integral whole, and means and ends as complementary. However, neither book cites the people's movement groups and academics who espouse such perspectives, and who warned early on of the limitations that are now being widely realized. Instead, the books attribute the positive influence to international funding agencies or to local non-governmental organizations, which have been influential in shaping the response in the first place. AIDS control has been led more by emotion, instinct and international politics than by scientific evidence or contextual needs. Denying this 'politics of knowledge' can only allow the suffering caused by AIDS to continue unabated.

*Combating AIDS* is slickly written using communicators' theories, taking the reader step-by-step through various arguments, and using repetition to ingrain them in the reader's mind. It is likely to interest local programme managers, as it provides examples of a wide range of interventions that comply with the homilies of current AIDS-control thinking. *Living with the AIDS Virus*, although less elegantly written, provides more food for thought for policy-makers, researchers and programme managers. Both books should interest the general reader. ■

Ritu Priya is at the Centre of Social Medicine and Community Health, Jawaharlal Nehru University, New Delhi 110067, India.

## Designer darwinism

**Darwin and Design: Does Evolution Have a Purpose?**

by Michael Ruse

Harvard University Press: 2003. 384 pp.

\$29.95, £19.95

Mark Ridley

Design — what biologists call 'adaptation' — is an obvious feature of life. People have probably been thinking about it for as long as they have been thinking about anything. Classically, it provided the basis of the 'argument from design', one of five arguments put forward for the existence of God. Darwin undermined that argument, but it has enjoyed something of a revival in the latest version of creationism, known as 'intelligent design creationism'.

Michael Ruse is a philosopher of biology, and his broad-ranging book covers several themes. He distinguishes the 'argument to complexity' — the factual observation that complex adaptations, such as eyes and wings, exist in nature — from the 'argument from design', the theological inference from that fact that God exists. Ruse traces the history of these two arguments from their earliest forms in the hands of Plato and Aristotle, through Galen and Thomas Aquinas, on to David Hume, Immanuel Kant and William Paley, followed by Georges Cuvier and Richard Owen, and then Darwin and his followers in the modern synthesis.

After the history, Ruse takes a look at modern research on adaptation, and at its 'formalist' critics, who think that organisms are shaped by non-adaptive laws of form, rather than by adaptation to the environment. He ends with two chapters on modern theologians, including the intelligent-design creationists.

There is no central argument to unify the book, but Ruse holds a consistently darwinian position against all its critics. Adaptation exists, he says; it matters; it is not explained by God; it is explained, and with exemplary scientific propriety, by natural selection; and it is a legitimate topic for scientific research. The various people who have argued otherwise are making various kinds of mistake.

The book is written for philosophers and theologians as well as scientists. Indeed, the book will help to broaden the minds of most scientific readers. I doubt whether anyone else has read up on quite the range of authors that Ruse has, and he writes about them clearly and non-technically.

Ruse also advances certain theses along the way that will particularly interest *Nature* readers, and I'll concentrate on them. One question concerns adaptationist research. The biologists who do this research often argue not only that they find adaptation

interesting, but that it is peculiarly important in biology. Successful adaptationist research is explanatory in a way that makes most mechanistic research look descriptive. And any big theory about life has to be able to explain adaptation.

On this topic, Ruse gives a sympathetic hearing to the critics who, he says, "would raise the objection that adaptation and the associated design metaphor have their roots in British natural theology of the pre-Darwinian era, and they would ask again why, in the secular world of the twenty-first century, we should be bound by this retro-thinking." This seems to be close to confusing the contexts of discovery and of justification. Adaptationism may have its roots in pre-darwinian anglican theology, but that does not make it false. German textual scholarship may have its roots in lutheran theology, but that does not reduce the value of the texts.

I think the strongest argument that the critics can make is not that adaptation is unimportant, just one problem among many, but that it has been solved. This may level the playing field for adaptationist and non-adaptationist research now, but it does not reduce the conceptual importance of adaptation for understanding life.

Research into adaptation often uses optimality models, or something like them. It might aim to understand the form of an eye, for instance, on the assumption that eye shape is optimized for forming images. One common interpretation of this research is that it aims to understand how, and not whether, organisms are adapted. Ruse disagrees. "It is fairer to say that a two-way process is at work here. We adopt a background assumption of adaptationism and then



**An eye for design? The complexity of a chameleon's eyes doesn't imply a role for God.**

L. ARNDT/NATUREPL.COM



we test specific models. Inasmuch as things do not work, then we worry more about our adaptationist background as well as our model."

I do not agree with Ruse on this point, but his position is undoubtedly defensible. The only place I had a deep disagreement was in the history of ideas, with what Ruse has to say about Motoo Kimura and the neutral theory of molecular evolution — the idea that most molecular evolution proceeds by random genetic drift. Kimura's ideas receive only a brief mention before being dismissed.

I favour the view that Kimura's idea led to something of a paradigm shift in the late 1980s. Most biologists who work on molecular evolution nowadays assume that they are mainly studying the effects of random drift. This contrasts with earlier biologists who worked with non-molecular characters. Kimura's ideas seriously challenged, or at least restricted the scope of, adaptationism. If adaptationism could carry on much as before, it was for reasons that needed further attention. So Kimura deserves more than a passing gesture in a survey as admirably thorough and sensible as *Darwin and Design*.

Mark Ridley is in the Department of Zoology, Oxford University, Oxford OX1 3RS, UK.

## Chemical conversations

### Candid Science III: More Conversations with Famous Chemists

by István Hargittai, edited by Magdolna Hargittai

World Scientific/Imperial College Press: 2003. 520 pp. £48, \$78 (hbk); £21, \$34 (pbk)

Arthur Greenberg

This book is the latest in a continuing series of interviews with prominent scientists by István and Magdolna Hargittai. The Hargittais are highly regarded structural chemists, located in Budapest, Hungary, whose wide-ranging interests and energy have produced a 'cottage industry' of books.

*Candid Science III* provides snapshots of the culture and progress of chemistry over the final 60 years of the twentieth century. Its oldest interviewee, Glenn Seaborg (1912–1999), was a nuclear chemist who added a new row — the actinides — to the periodic table. Seaborg was one of the team that discovered plutonium, and worked on the extraction of this metal for the Manhattan Project. His frustration with the naming of the trans-uranium elements is apparent in this interview from 1995. But in a postscript that was first published in 1998, Seaborg



István Hargittai has interviewed many of the most influential chemists of the past sixty years, including Glenn Seaborg (top left), Ad Bax (second row, right) and Jean-Marie Lehn (third row, second left).

happily describes the history of the discovery of element 106, which was named in his honour in 1997.

The youngest interviewee, born in 1956, is Ad Bax, who runs the US National Institutes of Health's section of NMR (nuclear magnetic resonance) spectroscopy. From 1981 to 1997, the ISI logged 21,655 citations of his work, 50% more than the second-most-cited chemist, Nobel laureate John A. Pople. This emphasizes the amazing evolution over the past half-century of NMR spectroscopy from a physics experiment to a routine technique for structural proof, and finally to a method for solving protein structures in solution and for magnetic imaging. Bax, who is at the cutting edge of this work, briefly describes the integration of NMR experiment and computation that took off during this 50-year period, due, in part, to unimaginable advances in computer technology. These applications of NMR, along with advances in X-ray crystallography (also referred to in the interviews with Johann Deisenhofer and Robert Huber), contributed mightily to the emerging field of structural biology, which arguably began with the discovery of the structure of DNA in 1953.

Reading this book reveals the growing sophistication with time of the interviewer's technique. For example, Hargittai began his 1996 interview with Nobel laureate Jean-Marie Lehn as follows: "You started as an organic chemist", to which Lehn replied: "Yes,

an organic chemist and, in fact, a natural products chemist." The interviewer's follow-up was: "Now you are also a very conceptual chemist. Not every organic chemist develops general concepts the way you do." Well, *bien sur*. Not many chemists of any stripe develop concepts like Lehn does.

In contrast, an interview two years later with another Nobel laureate, Bruce Merrifield, is more thoughtful and leads to some interesting insights. In discussing his early life in California during the Depression, Merrifield, who was seven years old when the stock market crashed, describes his family's economic condition: "So we scraped along, and that affected me all my life. I can't bear to waste things, I'm not extravagant, never buy anything on credit, and that came directly from the Depression." Two decades later, Merrifield developed solid-phase protein synthesis, a technique that essentially wastes none of the synthetic amino acids and peptides.

Most of the interviews are fairly straightforward affairs, but once in a while some sparks fly. For example, Hargittai starts the interview with Paul von Ragué Schleyer with a provocative implied question: "I recently heard you say that experiments are no longer necessary in chemistry. We can compute everything." Schleyer's response is forceful and there is a fascinating thrust and parry during the interview. The Schleyer interview tracks the career of this experimental chemist, whose early computations

## Science in culture

## The arsenic green

**William Morris seemed indifferent to the fact that his wallpapers contained arsenical pigments.**

Andy Meharg

William Morris (1834–1896) was a utopian idealist whose life was full of contradictions. He was a progenitor of the green movement and decried the environmental and human degradation caused by industrial activity. But he was also a successful capitalist who supplied the bourgeoisie with expensive interior décor. This paradox is most disturbingly evident in his intimate associations with arsenic.

His father, William Morris senior, helped set up the mining company Devon Great Consols (DGC), in its day the largest producer of arsenic in the world. William Morris used his income from shares in DGC to finance his design company Morris, Marshall, Faulkner & Co. (later Morris & Co.). He also served as a director of DGC between 1871 and 1875, when arsenic production was at its height. But he was increasingly influenced by the growing socialist movement and resigned his directorship in 1875.

The environmental pollution caused by DGC was vast and persists to this day. Working conditions were atrocious. Workers suffered widely from skin lesions known as arsenic 'pock', which caused great discomfort, and many died from arsenic-related lung disease.

From 1867 onwards, DGC was the major supplier of arsenic for the production of green pigments following the synthesis in the late eighteenth century of copper arsenite, named Scheele's green after its discoverer. These pigments were widely used in wallpapers. In damp rooms, fungi living on the wallpaper paste turned the arsenic salts into highly toxic trimethylarsine. Arsenic pigments, which were also used extensively in paints and to dye clothes, paper, cardboard, food, soap, and artificial and dried flowers, were responsible for untold numbers of cases of chronic illness and many deaths.

Was Morris using the arsenic from DGC in his own products? Evidence to suggest that he was comes from correspondence with Thomas Wardle, his dye manufacturer. In one letter, dated 3 October 1885, responding to an enquiry from a concerned customer (Mr Nicholson), Morris dismissed the concerns of the medical professions and popular press about arsenic poisoning wallpapers coloured

with arsenic pigments. "As to the arsenic scare, a greater folly is hardly possible to imagine: the doctors were being bitten by witch fever."

A second letter three days later stated: "Of course it is proving too much to prove that the Nicholsons were poisoned by wall-papers; for if they were a great number of people would be in the same plight and we should be sure to hear of it." This blasé response to the harmful effect that his wallpapers might have been having on his customers' health is remarkable, given the documented occurrence of arsenic poisoning at DGC.

To help me investigate the possible use of arsenic pigments in William Morris wallpapers, the William Morris Gallery in London kindly sent me a small piece of an early example of the 'Trellis' pattern wallpaper. The Trellis pattern is believed to be Morris's first wallpaper and was produced from 1864 onwards. I analysed the green pigment by energy-dispersive analysis and showed unequivocally that the coloration was caused by a copper arsenic salt. The beauty that William Morris wallpapers brought to a room must have had a health cost, at least in damp houses.

It is easy to be harsh on William Morris for his double standards with respect to industry and pollution. In his defence, he was a product of his age, when environmentalism was in its infancy. He was actually a positive force in this movement. His political creed developed over several decades, and by the end of his life, when he was most revolutionary, his links with industry were in the past.



**Hidden danger:**  
in damp houses, the green pigment  
in William Morris's wallpaper could  
have released toxic arsenic compounds.

However, as a writer who demonized the industrial practices of his time as dehumanizing, he is almost as silent as a stone on his own role in the most polluting of industries, arsenic production. The Wardle letters — all that survive of his thoughts on his connections with arsenic — show only indifference, not regret.

Andy Meharg is in the School of Biological Science, Cruikshank Building, Aberdeen University, Aberdeen AB24 3UU, UK.

helped to rationalize interesting results, and whose later work using much faster computers predicted utterly non-classical organometallics, some of which were later verified experimentally.

Some interviewees offer insights that are worthy of further discussion. Jacqueline K. Barton is self-reflective and generous to students and colleagues in tracing her career path. Starting on tenure track at Hunter College in New York, she moved to Columbia University, where she was promoted to full professor, and then moved to the Cali-

fornia Institute of Technology, where she holds an endowed professorship. She notes in passing that "the paths of woman professors tend to be non-traditional". Barton concludes the interview: "You can be a woman scientist in a major research institution and also be a person with a family and be happy." It is not a self-satisfied remark but rather active encouragement to young women pondering futures in science.

This book makes interesting light reading, especially for chemists who have watched the field develop over the past 30 to 60 years.

One might wish for a preface that provides some integration and perspective, for example in considering the career paths of the three women interviewed (Barton, Mildred Cohn and Reiko Kuroda). Nonetheless, the book makes a worthwhile contribution to the oral history of science, and I recommend it for both libraries and individuals.

Arthur Greenberg is professor of chemistry and dean of the College of Engineering and Physical Sciences, University of New Hampshire, Durham, New Hampshire 03824, USA. He is the author of *The Art Of Chemistry*.



# Molecular messages

Jack W. Szostak

In this age of genome sequencing, the idea that biopolymer sequences are a type of molecularly coded information is well established. We are all familiar with the idea that it is the sequence of the nucleotides or amino acids that make up DNA, RNA or protein molecules that determine their structure and function. But the recent deluge of phylogenetic sequence data provides thousands of examples of related but different sequences encoding essentially identical structures and functions. More radical are the accumulating examples of both RNA and protein molecules with entirely different structures but similar biochemical functions (for example, various structurally distinct protease enzymes have been identified). Such examples raise important questions about the nature of the information content of biological sequences. How best can we define and quantify the information content of biopolymer sequences?

The information content of biopolymers is usually thought of in terms of the amount of information required to specify a unique sequence or structure. This viewpoint derives from classical information theory, which does not consider the meaning of a message, defining the information content of a string of symbols as simply that required to specify, store or transmit the string. Thus, the unannotated human genome sequence can be encoded in a 750-megabyte file, but this could be greatly reduced in size by the application of standard data-compression techniques to account for internal repetitions.

Approaches such as algorithmic complexity further define the amount of information needed to specify sequences with internal order or structure, but fail to account for the redundancy inherent in the fact that many related sequences are structurally and functionally equivalent. This objection is dealt with by physical complexity, a rigorously defined measure

of the information content of such degenerate sequences, which is based on functional criteria and is measured by comparing alignable sequences that encode functionally equivalent structures. But different molecular structures may be functionally equivalent. A new measure of information — functional information — is required to account for all possible sequences that could potentially carry out an equivalent biochemical function, independent of the structure or mechanism used.

By analogy with classical information, functional information is simply  $-\log_2$  of the probability that a random sequence will encode a molecule with greater than any given degree of function. For RNA sequences of length  $n$ , that fraction could vary from  $4^{-n}$  if only a single sequence is active, to 1 if all sequences are active. The corresponding functional-information content would vary from  $2n$  (the amount needed to specify a given random RNA sequence) to 0 bits. As an example, the probability that a random RNA sequence of 70 nucleotides will bind ATP with micromolar affinity has been experimentally determined to be about  $10^{-11}$ . This corresponds to a functional-information content of about 37 bits, compared with 140 bits to specify a unique 70-mer sequence. If there are multiple sequences with a given activity, then the corresponding functional information will always be less than the amount of information required to specify any particular sequence. It is important to note that functional information is not a property of any one molecule, but of the ensemble of all possible sequences, ranked by activity.

Imagine a pile of DNA, RNA or protein molecules of all possible sequences, sorted by activity with the most active at the top. A horizontal plane through the pile indicates a given level of activity; as this rises, fewer sequences remain above it. The functional information required to specify that activity is  $-\log_2$  of the fraction of sequences above the plane. Expressing this fraction in terms of

information provides a straight-

forward, quantitative measure of the difficulty of a task. More information is required to specify molecules that carry out difficult tasks, such as high-affinity binding or the rapid catalysis of chemical reactions with high energy barriers, than is needed to specify weak binders or slow catalysts. But precisely how much more functional information is required to specify a given increase in activity is unknown. If the mechanisms involved in improving activity are similar over a wide range of activities, then power-law behaviour might be expected. Alternatively, if it becomes progressively harder to improve

## Functional information

*A quantitative means of comparing the functional abilities of different biopolymers would allow us to dissect out differences and to discern their origins.*

activity as activity increases, then exponential behaviour may be seen. An interesting question is whether the relationship between functional information and activity will be similar in many different systems, suggesting that common principles are at work, or whether each case will be unique.

The challenge in determining experimentally the relationship between functional information and activity is the extreme rarity of functional sequences in populations of random sequences (typically  $10^{-10}$  to  $10^{-15}$  for aptamers and ribozymes isolated from random RNA pools). *In vitro* selection and amplification allow the isolation of rare functional sequences from a large initial pool of random sequences. Unfortunately, the original distribution of functional molecules can be obscured by biases in replication and selection efficiency that accumulate over cycles of enrichment. A radically different approach would be to apply the new single-molecule fluorescence methods to the direct analysis of large sets of random sequences. Such experiments might ultimately allow us to understand why proteins have taken over so much of biochemical function from RNA, and they might also serve to guide and interpret the results of experiments in which new nucleotides or amino acids are used to expand the genetic code as we search for molecules even better than those supplied by nature. ■

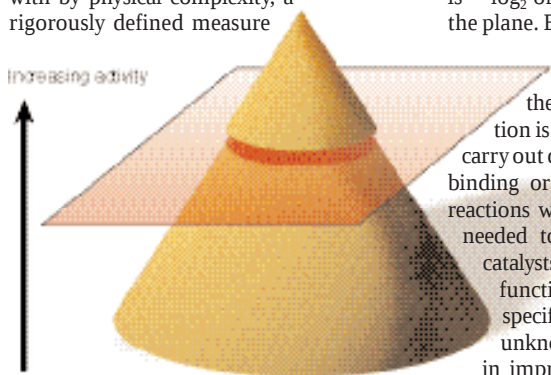
Jack W. Szostak is in the Howard Hughes Medical Institute and Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts 02114-2696, USA.

### FURTHER READING

Hamming, R. W. *Coding and Information Theory* (Prentice-Hall, Englewood Cliffs, New Jersey, 1987).  
Zurek, W. H. in *Studies in the Sciences of Complexity* Vol. 8 (ed. Pines, D.) (Addison-Wesley, Reading, Massachusetts, 1991).  
Adami, C. & Cerf, N. J. *Physica D* **137**, 62–69 (2000).  
Wilson, D. S. & Szostak, J. W. *Annu. Rev. Biochem.* **68**, 611–648 (1999).

### Erratum

In Antonio Damasio's Concepts essay on "Mental self" (*Nature* **423**, 227; 2003), the name of Gus Nossal was misspelt.



**Increasing activity implies fewer sequences and greater functional-information content.**



# A question of timing

Conel Alexander

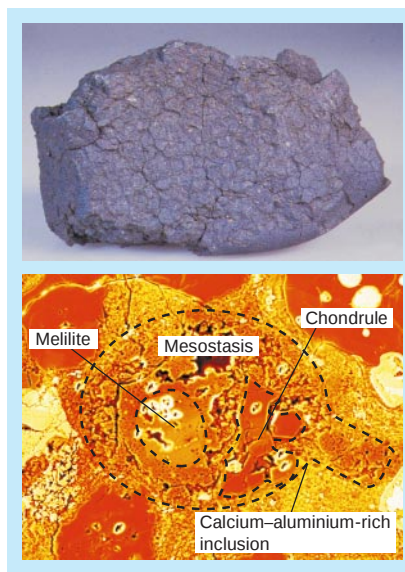
**Meteorites record the early history of the inner Solar System. A unique object that has been found in one meteorite may add support to a revolutionary idea about how the Solar System formed.**

Astronomers and astrophysicists have made great strides in identifying the basic stages in the formation of Sun-like stars and their planetary systems<sup>1</sup>. Stars form by the collapse of high-density regions, or cores, inside interstellar clouds. A disk of gas and dust forms around the growing star, and most of the mass that is accreted by the star passes through this disk. Early on, jets of partially ionized gas develop along the system's axis of rotation. The origin of these bipolar jets is hotly debated but, as we will see, they may have a lasting influence on any planets that form in the inner part of the disk.

Although this general picture is well established, very little is known about the precise conditions and processes that occur in a disk during these early stages. This is because it has not been possible to make direct observations of a disk's mid-plane region, where planets would form. So far, the best (albeit imperfect) record of what happened in one disk — the disk from which our Solar System formed — is found in meteorites<sup>2</sup>. On page 728 of this issue, Itoh and Yurimoto<sup>3</sup> describe an object discovered in a meteorite that they claim supports a revolutionary model of early Solar System evolution.

With a few exceptions, meteorites are fragments from the asteroid belt that lies between the orbits of Mars and Jupiter. Their parent asteroids are the last vestiges of the swarm of planetesimals from which the terrestrial planets — Mercury, Venus, Earth and Mars — formed. The oldest and most primitive meteorites, the chondrites, appear to consist largely of aggregates of material that formed in the disk before, or at the same time that, the planetesimals formed.

The major constituents of chondrites (50–80% by volume) are chondrules — small spheres of silicates, iron metal and iron sulphide. Up to a further 5% is made of predominantly silicate objects that are rich in calcium and aluminium, known as calcium–aluminium-rich inclusions, or CAIs. Chondrules and CAIs range in diameter from tens of micrometres to centimetres, and formed during brief heating episodes at temperatures of 1,600–2,000 K. The chondrules and some CAIs melted, and even partially vaporized; other CAIs appear to have condensed out of a hot gas and never melted. When they formed, they contained significant concentrations of short-lived



**Figure 1 Rock record.** The meteorite Y-81020 (top) — 140 mm across and weighing just over 270 g — may hold a vital clue to the early history of the Solar System. This false-colour image of a small section of the meteorite (bottom), taken by Itoh and Yurimoto<sup>3</sup> through electron back-scattering using a scanning electron microscope, shows what appears to be a calcium–aluminium-rich inclusion (CAI). The object, 100  $\mu\text{m}$  in diameter, has three components, which are roughly indicated by the dashed lines: a melilite crystal; a porous, fine-grained mesostasis; and, most interesting of all, a chondrule fragment. If this is indeed a CAI, the presence of a chondrule inside it implies that the chondrule formed before or contemporaneously with the CAI — contradicting models of the early Solar System in which CAIs formed up to two million years before chondrules.

radionuclides, although these concentrations are much lower in chondrules than in CAIs. The shortest-lived of these radionuclides ( $^{41}\text{Ca}$ ) has a half-life of only about 100,000 years, which means that the time interval between its synthesis and its incorporation into CAIs must have been, by astronomical standards, very short. A popular explanation for this short interval is that a nearby supernova both triggered the collapse of the fledgling Solar System and seeded it with the short-lived radionuclides.

Numerous mechanisms for the formation of chondrules have been proposed over

the years. One promising idea suggests that they formed in shock waves<sup>4</sup>, possibly generated by Jupiter, that repeatedly passed through the asteroid belt. How CAIs formed is even less certain. Equally troubling is that, if taken as being the result of differences in formation times, the higher abundances of short-lived radionuclides in CAIs imply that they formed between one and two million years before chondrules. Because, in general, disk material would then have been flowing into the growing Sun, explaining how CAIs could have been stored in the disk for this length of time is a considerable, but perhaps not an insurmountable, problem.

A radically different explanation for the existence of chondrules and CAIs, put forward by Shu *et al.*<sup>5</sup>, links their formation to the presence of bipolar jets in the early Solar System (Fig. 3 on page 730). These authors propose that bipolar jets were launched by strong magnetic fields at the interface between the disk and the forming star. The CAIs, they argue, formed in a region slightly closer to the Sun, where there was little gas, and were repeatedly irradiated by energetic flares emanating from the active young Sun. These flares not only melted or even vaporized the proto-CAIs, but also induced nuclear reactions that produced the short-lived radionuclides. Chondrules, Shu *et al.* suggest, formed at the same time in a gas-rich region slightly further from the Sun. Here the flares would have melted the proto-chondrules, but the shield of gas and dust would have reduced the amount of short-lived radionuclides that could be produced. The position of the launch point for the jets might have wandered through these two regions, sweeping up both chondrules and CAIs onto ballistic trajectories so that they rained down on much of the inner Solar System.

If Shu *et al.*<sup>5</sup> are correct, the abundance of chondrules and CAIs in chondrites implies that most material now in the inner Solar System was processed through the jets, strongly influencing the chemistry of the dust from which the terrestrial planets formed. It has the added attraction of requiring no mechanism to explain how CAIs were stored in the disk for up to two million years before the chondrules formed. However, as discussed in ref. 6, studies of the abundances of two short-lived radionuclides ( $^{10}\text{Be}$  and  $^{60}\text{Fe}$ ) in meteorites have recently called the

Shu *et al.* model into question, and a roughly million-year time difference between chondrule and CAI formation seems to have been confirmed by recent lead-isotope data.

But these arguments against the Shu *et al.* model would be greatly weakened if it were shown that chondrules and CAIs formed contemporaneously. Although rare, CAI fragments have been found within chondrules, and this is consistent with chondrules forming either in the same period as or after CAIs. Finding a chondrule fragment poor in short-lived radionuclides inside a CAI rich in short-lived radionuclides would be unambiguous evidence in support of the Shu *et al.* model. Itoh and Yurimoto<sup>3</sup> have found what they believe is just such an object.

In a section of the meteorite Y-81020, held at the National Institute of Polar Research, Tokyo, Itoh and Yurimoto have found what appears to be a CAI made up of three components: a chondrule fragment; a melilite (silicate) crystal that is probably a fragment of an earlier CAI; and a porous, fine-grained calcium–aluminium-rich silicate that cements the object together, the ‘mesostasis’ (Fig. 1). The mesostasis probably formed during the final melting that produced the object. Itoh and Yurimoto do not present any short-lived-radionuclide data, and so we must rely on inferences made from various physical and chemical features of the object to determine if it is what they claim. The inferred peak temperature (above 1,823 K) and cooling rate (between tens and hundreds of degrees per hour) experienced

by the object are broadly consistent with those of chondrules and CAIs. The oxygen-isotope compositions of the chondrule fragment and mesostasis are typical of chondrules and CAIs, respectively. But there are two puzzling features of the chondrule fragment. Its edges seem quite angular, suggesting that, unlike the melilite crystal, it was unaffected by melting of the mesostasis. It also contains the iron-sulphide mineral troilite. This volatile mineral is present in chondrules, but would not have been stable under the conditions of CAI formation. So it is surprising that it has survived.

Could this object be a chondrule whose precursors were dominated by much older CAI material, rather than a chondrule within a CAI? Arguments will certainly be made for both interpretations. Ultimately, the issue may only be resolved if the abundances of one or more of the short-lived radionuclides in the mesostasis can be determined. Much will be riding on these measurements. ■

Conel Alexander is in the Department of Terrestrial Magnetism, Carnegie Institution of Washington, 5241 Broad Branch Road, Washington DC 20015-1305, USA.  
e-mail: alexande@dtm.ciw.edu

1. O'Dell, C. R. & Beckwith, S. V. *Science* **276**, 1355–1358 (1997).
2. Alexander, C. M. O'D., Boss, A. P. & Carlson, R. W. *Science* **293**, 64–68 (2001).
3. Itoh, S. & Yurimoto, H. *Nature* **423**, 728–731 (2003).
4. Desch, S. J. & Connolly, H. C. Jr *Meteoritics Planet. Sci.* **37**, 183–208 (2002).
5. Shu, F. H., Shang, H., Gounelle, M., Glassgold, A. E. & Lee, T. *Astrophys. J.* **548**, 1029–1050 (2001).
6. Zinner, E. *Science* **300**, 265–267 (2003).

in fits and starts (punctuational evolution). Other questions concern the relationship between genetic, morphological and behavioural changes, and the precise region, or regions, of origin.

For instance, possible early *H. sapiens* fossils, dating from about 260,000 to 130,000 years ago, are scattered across Africa at sites such as Florisbad (South Africa), Ngaloba (Tanzania), Eliye Springs and Guomde (Kenya), Omo Kibish (Ethiopia), Singa (Sudan) and Jebel Irhoud (Morocco). But the best dated of these finds, from Florisbad and Singa, are problematic because of incompleteness and, in the latter case, evidence of disease. Meanwhile, the more complete or diagnostically modern specimens suffer from chronological uncertainties. So the most securely dated and complete early fossils that unequivocally share an anatomical pattern with today's *H. sapiens* are actually from Israel, rather than Africa. These are the partial skeletons from Skhul and Qafzeh, dating from around 115,000 years ago. Their presence in the Levant is usually explained by a range expansion from ancestral African populations, such as those sampled at Omo Kibish or Jebel Irhoud<sup>2,7,8</sup>, around 125,000 years ago.

The new cranial material from Herto, Ethiopia — described by White and colleagues<sup>4,5</sup> — adds significantly to our understanding of early *H. sapiens* evolution in Africa. The fossils are complete enough to show a suite of modern human characters, and are well constrained by argon-isotope dating to about 160,000 years ago. Three individuals are represented by separate fossils: a nearly complete adult cranium (skull parts excluding the lower jaw), a less complete juvenile cranium, and some robust cranial fragments from another adult<sup>4</sup>. All display evidence of human modification, such as cut marks, considered to represent mortuary practices rather than cannibalism. Associated layers of sediment produced evidence of the butchery of large mammals such as hippopotamuses and bovines, as well as assemblages of artefacts showing an interesting combination of Middle Stone Age and late Acheulean technology<sup>5</sup>.

The morphology of the most complete of these three fossils helps to clarify the pattern of early *H. sapiens* evolution in Africa, as it shows an interesting combination of features from archaic, early modern and recent humans. The cranium is very large, but once the size is standardized, it shares with ancient African crania a wide interorbital breadth (the distance between the orbits of the eyes), anteriorly placed teeth, and a short occipital (the bone at the rear of the braincase). It also has a wide upper face and moderately domed forehead, as do the Skhul and Qafzeh fossils. Its low nose and face and flat midface are more widely shared early *H. sapiens* features, whereas other

## Human evolution

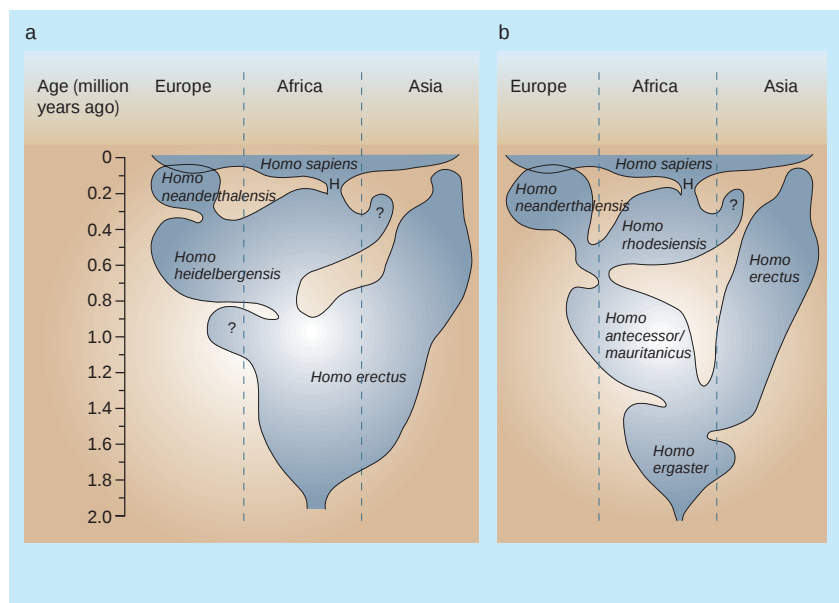
# Out of Ethiopia

Chris Stringer

Newly discovered fossils from Ethiopia provide fresh evidence for the 'out of Africa' model for the origin of modern humans, and raise new questions about the precise pattern of human evolution.

The idea that modern humans originated in Africa, with populations subsequently spreading outwards from there, has continued to gain support lately. But much of that support has come from analyses of genetic variation in people today<sup>1</sup>, and from fossil and archaeological discoveries dated to within the past 120,000 years<sup>2,3</sup> — after our species evolved. Hard evidence for the inferred African origin of modern humans has remained somewhat elusive, with relevant material being fragmentary, morphologically ambiguous or uncertainly dated. So the fossilized partial skulls from Ethiopia that are described on pages 742 and 747 of this issue<sup>4,5</sup> are probably some of the most significant discoveries of early *Homo sapiens* so far, owing to their completeness and well-established antiquity of about 160,000 years.

There are two broad theories about the origins of *H. sapiens*. A few researchers still support a version of the 'multiregional' hypothesis, arguing that the anatomical features of modern humans arose in geographically widespread hominid populations throughout the Pleistocene epoch (which lasted from around 1.8 million to some 12,000 years ago)<sup>6</sup>. But most now espouse a version of the 'out of Africa' model, although there are differences of opinion over the complexity of the processes of origin and dispersal, and over the amount of mixing that might subsequently have occurred with archaic (non-modern) humans outside of Africa<sup>2,7</sup>. Within Africa, uncertainties still surround the mode of modern human evolution — whether it proceeded in a gradual and steady manner or



**Figure 1** Origin of our species. The figure shows the geographical and temporal distribution of hominid populations, based on fossil finds, using different taxonomic schemes. The new finds from Herto<sup>4,5</sup> (H) represent early *Homo sapiens*. a, This reflects the view that both Neanderthals and modern humans derived from a widespread ancestral species called *H. heidelbergensis*<sup>2</sup>. b, However, evidence is growing that Neanderthal features have deep roots in Europe<sup>2,8</sup>, so *H. neanderthalensis* might extend back over 400,000 years. The roots of *H. sapiens* might be similarly deep in Africa, but this figure represents the alternative view that the ancestor was a separate African species called *H. rhodesiensis*. Different views of early human evolution are also shown. Some workers prefer to lump the earlier records together and recognize only one widespread species, *H. erectus*<sup>2</sup> (shown in a). Others recognize several species, with *H. ergaster* and *H. antecessor* (or *H. mauritanicus*) in the West, and *H. erectus* only in the Far East<sup>8</sup> (shown in b). Adapted with permission from refs 8, 11.

characteristics, such as its globular braincase, are typically modern. In the angulation and transverse ridge of the occipital, there is also an intriguing resemblance to fossils from sites such as Elandsfontein (South Africa) and Broken Hill (Zambia) that are often assigned to *H. heidelbergensis* or *H. rhodesiensis*. This may provide a clue to the individual's ancestors (Fig. 1). But overall, the fossil seems closest in morphology to particular crania from Jebel Irhoud, Omo Kibish and Qafzeh.

So White and colleagues' findings<sup>4,5</sup> provide a plausible link back to more ancient African fossils, and forward to Levantine samples. They also raise questions about the overall pattern of modern human origins in Africa. Because of Africa's great area and still limited fossil record, it is uncertain whether the pattern of *H. sapiens* evolution there was essentially continent-wide, or was a more localized — and perhaps punctuational — process. The Herto finds shift the focus once again to East Africa. It seems from these crania and from possibly contemporaneous fossils, such as those at Ngaloba, Singa and Eliye Springs, that human populations of this era showed a great deal of anatomical variation. So, did the early modern morphology spread outwards from East Africa, perhaps gradually

more archaic forms? Or could there have been an African version of multiregionalism, with modern morphology coalescing from various populations across the continent<sup>2,7,8</sup>? Only better samples and better dating of the African fossil record will help resolve these questions.

And what of the taxonomic status of the new finds? White and colleagues propose that, although measurements of the most complete fossil differentiate it from geologically 'recent' (that is, post-Pleistocene) *H. sapiens*, there is sufficient evidence to assign the material to this species overall, while naming a new subspecies, *idaltu*. However, in my opinion, the distinctive features described for *H. sapiens idaltu* might not be so unusual, and could probably be found in late Pleistocene samples from regions such as Australasia<sup>9</sup>.

Do the Herto fossils represent 'modern' *H. sapiens*? There is an ongoing debate about the concept of modernity, in terms of both morphological and behavioural characteristics<sup>2,3,7,8,10</sup>. Nevertheless, despite the presence of some primitive features, there seems to be enough morphological evidence to regard the Herto material as the oldest definite record of what we currently think of as modern *H. sapiens*. The fact that the geological age of these fossils is close to some estimates



#### 100 YEARS AGO

It is reported that a young Austrian doctor named Sachs has fallen a victim to his scientific zeal having accidentally inoculated himself with plague, from the effects of which he died after a short illness. Such regrettable incidents will occur while scientific research is pursued, and cannot be avoided even by the greatest foresight. There is no likelihood that other cases will develop, as under good hygienic conditions plague is not particularly infectious from man to man, and European doctors and nurses tending the sick seldom contract the disease.

#### ALSO...

In the course of a recent article published in the *Recueil de l'Institut botanique de Bruxelles*, Prof. Errera comes to the conclusion that it is not possible for organisms to exist of a size very appreciably smaller than those which can be observed with the highest powers of the microscope now in use. An estimation is made of the number of molecules of certain bodies, such as albuminoids, which are present in a bacterium of given size: the number is of such an order of magnitude that only a few molecules could be present in an organism having a diameter  $0.01\mu$ , and thus a minimum limit to the possible size is obtained.

From *Nature* 11 June 1903.

#### 50 YEARS AGO

Meeting on "Preservation of Normal Tissues for Transplantation". In opening the scientific proceedings, Prof. P. B. Medawar (University College London) said that living skin, when transplanted into positions formerly occupied by skin, was probably the most exacting of all tissues... Under the heading [of modifying host reactions] Prof. Medawar outlined experiments done in collaboration with R. E. Billingham and L. Brent which showed that if an animal were presented with living foreign cells in foetal life, its power to react against those cells in later life was reduced or wholly abolished. This was not due, as had been widely assumed, to an adaptation of the grafted cells, but to an adaptation of the host, for 'actively acquired tolerance', once established by inoculation of the foetus, extended to cells freshly transplanted in later life — cells which therefore had had no opportunity to adapt themselves to alien soil.

From *Nature* 13 June 1953.



obtained by genetic analyses for the origin of modern human variation<sup>1</sup> only heightens their importance.

Chris Stringer is in the Human Origins Group at The Natural History Museum, London SW7 5BD, UK.

e-mail: c.stringer@nhm.ac.uk

1. Ingman, M., Kaessmann, H., Pääbo, S. & Gyllenstein, U. *Nature* 408, 708–713 (2000).
2. Stringer, C. *Phil. Trans. R. Soc. Lond. B* 357, 563–579 (2002).
3. McBrearty, S. & Brooks, A. *J. Hum. Evol.* 39, 453–563 (2000).
4. White, T. D. *et al. Nature* 423, 742–747 (2003).

5. Clark, J. D. *et al. Nature* 423, 747–752 (2003).
6. Wolpoff, M. & Caspari, R. *Race and Human Evolution: A Fatal Attraction* (Simon & Schuster, New York, 1997).
7. Lahr, M. & Foley, R. *Yb. Phys. Anthropol.* 41, 137–176 (1998).
8. Hublin, J.-J. in *Human Roots: Africa and Asia in the Middle Pleistocene* (eds Barham, L. & Robson-Brown, K.) 99–121 (Western Academic & Specialist Press, Bristol, 2001).
9. www-personal.une.edu.au/~pbrown3/palaeo.html
10. Lieberman, D., McBratney, B. M. & Krovitz, G. *Proc. Natl Acad. Sci. USA* 99, 1134–1139 (2002).
11. Rightmire, G. P. in *Human Roots: Africa and Asia in the Middle Pleistocene* (eds Barham, L. & Robson-Brown, K.) 123–133 (Western Academic & Specialist Press, Bristol, 2001).

## High-energy physics

# Into the fifth dimension

Juan Maldacena

Particles such as the proton can be imagined as vibrating strings. We also know that protons contain smaller, point-like particles, going against the string theory. But in five dimensions, the contradiction disappears.

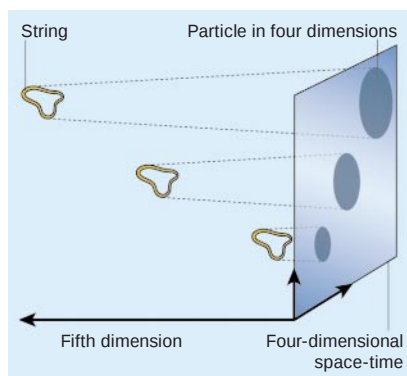
In fundamental physics, our description of nature involves four forces: gravitational, electromagnetic, weak and strong. The strong force is responsible for binding protons and neutrons inside the atomic nucleus. Two different theoretical approaches have been taken in describing the workings of the strong force and the structure of particles such as the proton and neutron. The theories are seemingly at odds with each other, but steps are gradually being taken to reconcile the two. Writing in the *Journal of High Energy Physics*, Polchinski and Strassler<sup>1</sup> now dispel worries over an apparent contradiction between the theories, by showing that it isn't necessarily a contradiction at all.

In the 1960s, experiments on high-energy collisions between protons revealed a plethora of other short-lived, strongly interacting particles. Shortly afterwards, a theory emerged that proposed that all of these different particles were particular excitation modes of a string: as a violin string can vibrate with different frequencies, these strings could oscillate in different ways, corresponding to the 'zoo' of particles that was observed. This 'string theory' proved useful in explaining some aspects of the masses and spins of the particles.

But further experiments carried out through the 1970s showed that protons are not fundamental particles. In the same way that, much earlier in the century, Rutherford had shown that the atomic nucleus was much smaller than an atom, experimenters showed that protons, and neutrons, have small point-like constituents. This didn't fit with the theory of protons as strings, which are extended objects. In fact, these experiments led to a new description of the strong interaction in terms of point-like quarks and gluons, through a theory called quantum chromodynamics (QCD).

As the electron carries an electric charge,

quarks and gluons carry a new type of charge, called 'colour' (hence 'chromodynamics'). The gluons transmit the strong force between quarks in much the same way that the photon transmits the electromagnetic force between electrons and other charged particles. To describe the strong force we need three 'colours' — three different types of charges, usually designated 'red', 'green' and 'blue'. The validity of QCD has been spectacularly confirmed by experiments at high energies in particle colliders. But, despite this success, it is still remarkably hard to do theoretical calculations with QCD at low energies. And that's exactly where things should get interesting: at low energies, the colour flux lines form bundles of energy



**Figure 1 Strings, particles and extra dimensions.** Strings moving in the fifth dimension are represented in the everyday world by their projection onto the four-dimensional boundary of the five-dimensional space-time. The same string located at different positions along the fifth dimension corresponds to particles of different sizes in four dimensions: the further away the string, the larger the particle. The projection of a string that is very close to the boundary of the four-dimensional world can appear to be a point-like particle.

that should behave like a string — a tantalizing connection from QCD to string theory. These strings, made of gluons, bind the quarks together.

In fact, in the 1970s, Gerard 't Hooft<sup>2</sup> showed that QCD becomes a theory of free (non-interacting) strings if the number of colours is infinite. This simplifies the theory considerably. Strings still exist in the three-colour version of QCD, but in this case the strings are interacting. No way has yet been found to simplify QCD into a free-string theory, but it could be the key to understanding many low-energy properties of particles that interact through the strong force, and in particular for deriving a curious property of QCD, called confinement. No one has ever observed a free quark, because colour-charge-bearing objects such as quarks and gluons are subject to confinement: in other words, as two quarks are gradually separated the attractive force between them due to their colour charges remains constant; this contrasts with the more familiar forces in electromagnetism and gravity that fall off with the square of increasing distance.

The way forward has been signalled by work on strings in 'QCD-like' theories<sup>3–5</sup>. A surprising and counterintuitive feature of these strings is that they move in more than the familiar four dimensions of everyday life — three spatial dimensions and one of time. Even though the gluons that make up the strings move in four dimensions, the string itself moves in five dimensions. Polchinski and Strassler<sup>1</sup> now show that this fact is a crucial element in reconciling the string picture and the point-like behaviour seen in high-energy collisions.

The strings move in a five-dimensional curved space-time with a boundary. The boundary corresponds to the usual four dimensions, and the fifth dimension describes the motion away from this boundary into the interior of the curved space-time. In this five-dimensional space-time, there is a strong gravitational field pulling objects away from the boundary, and as a result time flows more slowly far away from the boundary than close to it. This also implies that an object that has a fixed proper size in the interior can appear to have a different size when viewed from the boundary (Fig. 1). Strings existing in the five-dimensional space-time can even look point-like when they are close to the boundary. Polchinski and Strassler<sup>1</sup> show that when an energetic four-dimensional particle (such as an electron) is scattered from these strings (describing protons), the main contribution comes from a string that is close to the boundary and it is therefore seen as a point-like object. So a string-like interpretation of a proton is not at odds with the observation that there are point-like objects inside it.

Because the theory that describes the interior of the five-dimensional space-time

includes gravity, there are other interesting consequences of this line of argument. One of the most striking is the following. In QCD, when the temperature reaches sufficiently high values (above  $10^{12}$  K), a phase transition occurs and quarks and gluons are no longer confined — instead, a ‘soup’ of free particles is formed, called quark–gluon plasma. In the five-dimensional theory, this transition also corresponds to the formation of a black hole in the interior. Our knowledge of black holes can then tell us something about quark–gluon plasma. In addition, the QCD–string theory provides a simple explanation for an interesting feature of black holes — the Bekenstein–Hawking entropy. This entropy, a measure of the number of possible quantum microstates, arises from the thermodynamic properties of a black hole (which are also at the root of Hawking radiation). Counting these microstates to work out the entropy has proved a major challenge in the-

ories of quantum gravity. But in the five-dimensional theory, the black-hole entropy becomes just the entropy of the plasma of quarks and gluons.

There is an intimate connection between the physics of strong interactions and both string theory and quantum gravity. Hopefully, in the next few years a string-theory description for real-world QCD will emerge, making it possible to perform computations in a relatively simple way. And perhaps, beyond that, we might even arrive at a QCD-like theory that can describe gravity. ■

Juan Maldacena is in the School of Natural Sciences, Institute for Advanced Study, Princeton, New Jersey 08540, USA.  
e-mail: malda@ias.edu

1. Polchinski, J. & Strassler, M. J. *High Energy Phys.* 05(2003)012.
2. 't Hooft, G. *Nucl. Phys. B* **72**, 461–473 (1974).
3. Maldacena, J. *Adv. Theor. Math. Phys.* **2**, 231–252 (1998).
4. Gubser, S., Klebanov, I. & Polyakov, A. *Phys. Lett. B* **428**, 105–114 (1998).
5. Witten, E. *Adv. Theor. Math. Phys.* **2**, 253–291 (1998).

## Cognitive neuroscience

# Practice doesn't make perfect

Wilson Geisler and Richard Murray

It may seem counterintuitive, but we are not very efficient at recognizing even the most common words. This finding suggests strict limits on how flexible we are in learning to recognize new patterns.

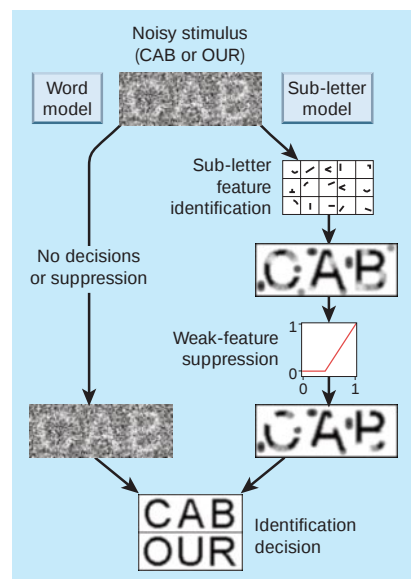
If there is any truth to the adage that ‘practice makes perfect’, we should certainly be near perfect at recognizing common words. As young children, we are taught first to recognize letters, and shortly thereafter words, and then for the rest of our lives we practise word recognition every day — the average literate person has read, at a conservative estimate, a hundred million words by the age of 25, thereby practising the recognition of each common word many hundreds of thousands of times. Nonetheless, on page 752 of this issue, Pelli, Farell and Moore<sup>1</sup> show that word recognition is surprisingly inefficient relative to letter recognition. This finding, which is surely not unique to words, implies that the neural learning mechanisms that are involved in pattern and object recognition are severely limited in their capabilities.

To measure the efficiency of word recognition, Pelli *et al.*<sup>1</sup> used the powerful tools and logic of ideal-observer analysis<sup>2–6</sup>. At the heart of this method is the concept of the ‘ideal observer’ — a theoretical device that achieves the best-possible performance in a perceptual or cognitive task, given the available information and constraints. Perceptual and cognitive tasks generally involve processing noisy physical or neural signals under conditions of uncertainty, and so ideal observers are typically derived using the concepts and methods of Bayesian statistics.

Pelli and colleagues’ study elegantly demonstrates the value of this type of analysis. First, derivation of the ideal observer forces one to rigorously specify the task at hand and to determine at least one specific mechanism for achieving optimal performance. Here, the task was to identify words or letters embedded in spatial white noise (top of Fig. 1). In each trial, a word was randomly picked from a set of possible words, given a particular contrast, and then added to a random sample of noise; the contrast determined how well the word stood out against the noisy background.

One type of mechanism that achieves optimal word recognition in this task is the ‘template matcher’ (Fig. 1, left path). This mechanism stores an exact copy (a template) of each target (such as a word) that could appear in a stimulus. To identify a particular target, the template matcher calculates how well each point on the target corresponds with each point on each stored template, and identifies the target as the template that gives the highest correlation<sup>7</sup>. (If the various targets do not occur with equal probability, then each correlation is adjusted according to the target’s probability.) Many neurons that are found in the first stages of the visual pathway can be modelled, to a first approximation, as template matchers.

The second benefit of ideal-observer analysis is that ‘ideal’ performance provides



**Figure 1** Why are we so inefficient at recognizing words? An efficient word-recognition mechanism, such as a template matcher (left pathway), might integrate information over the entire stimulus. As more and more letters are added to a word, the template matcher can identify the word at lower and lower contrasts. On the other hand, a mechanism that suppresses weak neural responses to letters or parts of letters (sub-letter features; right pathway) sets a lower limit on the contrast at which words can be identified: below the suppression threshold, the stimulus is simply not visible. This threshold is represented in the graph by the point at which the curve relating input (x-axis) to output (y-axis) begins to increase. Consequently, the addition of more and more letters does not mean that words can be identified at lower and lower contrasts. Pelli *et al.*<sup>1</sup> have found that our ability to recognize words decreases rapidly as word length increases, implying that the brain does not work as a ‘template matcher’. They argue that instead it functions more like the right pathway.

an appropriate benchmark against which to compare human performance. In Pelli and colleagues’ work<sup>1</sup>, it makes it possible to rigorously answer the question of whether practice makes perfect, by providing a precise description of ‘perfect’ performance. The performance of a human observer relative to the ideal is called the efficiency. In this case, the efficiency equals the squared contrast of the target at which it can be identified by the ideal observer with some given level of accuracy (for instance, 70% correct), divided by the squared contrast required for the human observer to identify it with the same level of accuracy. Pelli and colleagues’ main finding is that human observers’ efficiency at recognizing words declines drastically as word length increases. Specifically, if a person’s efficiency at recognizing single letters is  $F$ , then their efficiency at recognizing

$n$ -letter words is  $F/n$ . So, for example, common five-letter words are identified with only 1/5 the efficiency of single letters. This relatively poor performance holds for even the most common three-letter words, which often occur many times in a paragraph.

Why are people so inefficient at recognizing even common words? The ideal observer can also be helpful in addressing such questions, by providing a starting point for formulating possible models of human (sub-optimal) performance. We know that, with a little practice, humans can become extremely efficient at recognizing some simple patterns — such as spatially localized bars, edges and spots (see, for example, ref. 4). This is probably possible because the visual cortex in the primate brain contains neurons that act like template matchers for these simple patterns. But the  $1/n$  efficiency relationship reported by Pelli *et al.* implies that the brain cannot learn optimal templates for whole words. Rather, their results are quantitatively consistent with a model in which the brain has efficient templates only for parts of words — letters, or even features of letters — and in which only the strong responses of the neurons encoding these parts are passed on to word-identification processes (Fig. 1, right path). The hypothesis that familiar complex objects are initially encoded in smaller parts is consistent with recent psychophysical studies of the identification of objects in noise<sup>8</sup>.

Although Pelli and colleagues' results<sup>1</sup> show that word recognition is relatively inefficient, they do not rule out the possibility of specific neural circuits for the recognition of common words or other familiar objects. For example, there could be individual neurons whose receptive fields are precisely tuned to such objects. What the results do imply is that any neural circuits for recognizing complex objects such as words (or letters) cannot integrate neuronal responses as efficiently as can, say, the circuits underlying the receptive fields of neurons in the primary visual cortex (an early stage of the visual pathway), which encode simple patterns such as bars, edges and spots.

Pelli and colleagues suggest that inefficient integration occurs because weak responses to features (or letters) are suppressed ('squelched'). Moreover, they suggest that the brain makes discrete (categorical) decisions about which features (or letters) are present before deciding which word is present. Although this is plausible, there are alternative explanations that are similar in spirit but differ in detail. For example, it might not be necessary to invoke the suppression of weak responses to features. The  $1/n$  efficiency relationship could also be a result of increasing noise (neural noise, for instance) or uncertainty at lower target contrasts. The essential thing is that the signal-to-noise ratios of the neuronal responses

that are passed on to the word-identification process degrade rapidly as the target's contrast is lowered. Similarly, discrete decisions about features or letters are not required; simulations that we have carried out suggest that the predictions are almost the same if graded responses are used in word identification, as long as the graded responses deteriorate at low contrasts.

Regardless of the specific explanation, Pelli and colleagues' results<sup>1</sup> clearly show that, even with extensive practice over many years, perceptual learning mechanisms are unable to achieve efficient recognition of arbitrary objects. This finding is particularly revealing given that the visual system is very efficient at recognizing certain simple patterns such as bars and edges. It seems that Pelli *et al.* have uncovered, with the help

of ideal-observer analysis, a fundamental limit on the neural mechanisms of learning and plasticity. ■

Wilson Geisler and Richard Murray are at the Center for Perceptual Systems and the Department of Psychology, University of Texas, Austin, Texas 78712, USA.

e-mail: geisler@psy.utexas.edu

1. Pelli, D. G., Farell, B. & Moore, D. C. *Nature* 423, 752–756 (2003).
2. Green, D. M. & Swets, J. A. *Signal Detection Theory and Psychophysics* (Wiley, New York, 1966).
3. Barlow, H. B. *Vision Res.* 18, 637–650 (1978).
4. Burgess, A. E., Wagner, R. F., Jennings, R. J. & Barlow, H. B. *Science* 214, 93–94 (1981).
5. Geisler, W. S. *Psychol. Rev.* 96, 267–314 (1989).
6. Geisler, W. S. & Kersten, D. *Nature Neurosci.* 5, 508–510 (2002).
7. Peterson, W. W., Birdsall, T. G. & Fox, W. C. *Trans. Inst. Radio Engineers Prof. Group Information Theory* 4, 171–212 (1954).
8. Gold, J. M., Murray, R. F., Bennett, P. J. & Sekuler, A. B. *Curr. Biol.* 10, 663–666 (2000).

## Astrophysics

# Superficial resonance

Frits Paerels

Neutron stars are the most poorly understood stellar objects in the Universe. But observations of X-rays emitted from one neutron star have now revealed a clue to the nature of its surface and composition.

Neutron stars are the remnants of massive stars. The density of their cores exceeds that of atomic nuclei, and they might contain exotic states of matter that are not found anywhere else in nature<sup>1</sup>. For decades astronomers have tried to 'see' the surfaces of such stars, in an effort to determine their physical properties. Space-borne telescopes have brought improved X-ray-imaging capabilities, but the first observations of neutron stars with these instruments were rather disappointing, revealing that the X-ray emission from most isolated neutron stars is remarkably featureless. But now observations of one neutron star have finally given us a glimpse of the fundamental properties of these objects: on page 725 of this issue, Bignami *et al.*<sup>2</sup> describe the first physical clue to the nature of this star's X-ray emission.

When massive stars — stars more than eight times heavier than the Sun — run out of thermonuclear fuel, their cores collapse. The resulting configuration typically has a mass about one-and-a-half times that of the Sun, but a radius of only 10 km: the density of the material exceeds the nuclear density, and degeneracy (the quantum-mechanical restriction on the occupation of energy levels by particles) provides the support against the gravity of the star's own weight. According to simple astrophysical arguments, a young neutron star is expected to have a magnetic field of the order of  $10^{12}$  gauss at its surface (in comparison, Earth's field is about 0.5 gauss). It should also spin rapidly, appearing as a 'pulsar'.

The first neutron stars were discovered through their emission at radio wavelengths. Radio emission is produced in the envelope of magnetic field that surrounds a pulsar (through processes that are still poorly understood), and although it carries information about the star's magnetic field and spin period, it tells us nothing about the internal properties of the neutron star itself. If, instead, we could observe emission from the star's surface, it would tell us about the surface composition, which in turn would reveal more about how a neutron star forms, and the processes that affect the chemical composition of its atmosphere.

To 'view' the surface of hot young neutron stars requires observations of their X-ray, rather than radio, emission: the emission peaks in the 'soft' X-ray band of photon energies (0.1–1 kiloelectronvolts, or keV). But even in the X-ray band, isolated hot neutron stars are faint. The 1999 launches of NASA's Chandra X-ray Observatory and the European Space Agency's XMM-Newton observatory were eagerly awaited, but the initial observations of a handful of hot neutron stars were universally disappointing. The spectra of radiation from the stars did show the expected characteristic of thermal emission from their surfaces: a gently curving, black-body-like energy distribution. But detailed scrutiny of the X-ray spectrum failed to reveal any structure; the shape of the spectrum in all cases was very close to that of a black body, which by its very nature



contains no other information than the thermodynamic temperature<sup>3</sup>.

But last year Sanwal *et al.*<sup>4</sup> reported observations using Chandra of the spectrum of the young isolated neutron star 1E1207.4–5209. The data show two distinct dips in the energy spectrum of photons emitted from the star, centred at 0.7 and 1.4 keV. Here was the first spectral signature of an interaction of the surface radiation with something in the stellar atmosphere — but what? These features, known as absorption lines, were difficult to explain. Sanwal *et al.* speculated that they are due to absorption by singly ionized helium atoms in the stellar atmosphere; others, however, argued that transitions in highly ionized oxygen or neon atoms are the more likely explanation<sup>5,6</sup>.

The energies at which the features appear — 0.7 and 1.4 keV — are in the ratio 1:2. That suggests another interpretation. Charged particles in a magnetic field oscillate at a particular, resonant frequency — this is called ‘cyclotron resonance’. For a magnetic field strength of about  $6 \times 10^{10}$  gauss, the resonant frequency of electron oscillations corresponds to an absorbed-photon energy of 0.7 keV; absorption will also occur at multiples of this fundamental energy. Protons can resonantly absorb radiation in a magnetic field, but, for a given resonant frequency, the corresponding magnetic field would be 2,000 times higher than in the case of cyclotron resonance of electrons (because a proton is 2,000 times heavier than an electron).

From an accurate timing analysis of the spin behaviour of 1E1207.4–5209, Sanwal *et al.*<sup>4</sup> estimate that the star has a surface magnetic field strength of  $2\text{--}3 \times 10^{12}$  gauss. But if the 0.7-keV feature in the X-rays spectrum is a cyclotron resonance, this doesn’t add up — for an electron resonance, the field strength would be much lower than this value, whereas a proton resonance would require a field strength that is much higher. For this reason, the cyclotron-resonance interpretation was initially rejected.

But Bignami *et al.*<sup>2</sup> now report more detailed observations of the same neutron star using cameras onboard XMM-Newton, data that provide convincing proof of the cyclotron interpretation. Bignami *et al.* have found a third feature at about 2.1 keV and perhaps even a fourth feature around 2.8 keV — almost precisely three and four times 0.7 keV, respectively. What makes the identification compelling is that all of the features vary systematically in strength (and perhaps even in shape) with the spin phase of the star, ruling out the possibility of this being a chance conspiracy of noise fluctuations in the data. Cyclotron resonances have been seen before in a handful of neutron stars in binary systems, the first one being Hercules X-1 (ref. 7). But never before has a triple harmonic structure been seen so convincingly in an isolated neutron star.

The implications of this beautiful spectrum are somewhat puzzling, though. First, there is the obvious question of why only this object shows cyclotron resonances. (There is a recent report<sup>8</sup> of a broad feature in the spectrum of the isolated neutron star RBS1223, but even if that also turns out to be a cyclotron resonance, it still means that only a small fraction of isolated neutron stars show these kinds of spectral features.) This question translates into the equivalent statement that only a small fraction of isolated young neutron stars have a low magnetic field strength, one that is in the right range for electron cyclotron resonances to show up in the 0.1–10-keV band.

And there is the problem that the magnetic field strength estimated from the resonance energy is inconsistent with the field strength derived from the star’s observed spin rate and deceleration rate. Perhaps there are additional torques on 1E1207.4–5209. That would clearly have implications for our understanding of the angular momentum and magnetic field strength existing at the birth of a neutron star — an understanding that is already under siege on another flank, with the discovery of ‘magnetars’. These neutron stars have superstrong magnetic fields<sup>9</sup> that exceed a staggering  $10^{14}$  gauss.

Whatever the answer, the fact that 1E1207.4–5209 shows definite structure in

the spectrum of light emitted near or at its surface is an important finding. Physical clues to the fundamental properties of neutron stars are rare, and this discovery adds an important piece to the puzzle. It will be interesting to see whether the detailed shape of the spectrum can be made to fit a model for the atmosphere, and what the resulting distribution of the field, and perhaps the thermal structure of the atmosphere, looks like. And the discovery itself will undoubtedly add impetus to attempts to find spectral structure in the emissions of other neutron stars. ■

Frits Paerels is in the Department of Astronomy and the Columbia Astrophysics Laboratory, Columbia University, New York, New York 10027, USA.  
e-mail: frits@astro.columbia.edu

1. Miller, C. *Nature* **420**, 31–33 (2002).
2. Bignami, G. F., Caraveo, P. A., De Luca, A., & Mereghetti, S. *Nature* **423**, 725–727 (2003).
3. Zavlin, V. E. & Pavlov, G. G. in *Proc 270. WE-Heraeus Seminar on Neutron Stars, Pulsars, and Supernova Remnants* (eds Becker, W., Lesch, H. & Trümper, J.) 263–272 (Max Planck Inst. Extraterrestr. Phys., Garching, 2002).
4. Sanwal, D., Pavlov, G. G., Zavlin, V. E. & Teter, M. A. *Astrophys. J.* **574**, L61–L64 (2002).
5. Hailey, C. J. & Mori, K. *Astrophys. J.* **578**, L133–L136 (2002).
6. Mori, K. & Hailey, C. J. Preprint astro-ph/0301161 at <http://arXiv.org> (2003).
7. Trümper, J. *et al.* *Astrophys. J.* **219**, L105–L110 (1978).
8. Haberl, F., Schwöpe, A. D., Hambaryan, V., Hasinger, G. & Motch, C. *Astron. Astrophys.* **403**, L19–L23 (2003).
9. Duncan, R. C. & Thompson, C. *Astrophys. J.* **392**, L9–L13 (1992).

## Evolutionary biology

# Genes to make new species

Mohamed A. F. Noor

A long-term goal of studies of the way in which new species form has been to identify the genes involved, and the forces that drive their evolution. That goal is now being realized — and natural selection plays a major part.

For some 70 years, researchers have been crossing different fruitfly species in an attempt to answer one of the most fundamental questions in evolutionary biology: what are the genetic changes that cause one species to split into two? Speciation is known to involve genes that prevent two individuals from mating — or, if they do mate, from producing living, fertile offspring. These genes contribute, for instance, to mate discrimination, the success of fertilization, adaptation to particular environments, and the physiological defects of the ‘hybrid’ offspring that are produced when two species do interbreed. But the identity of such genes, their normal functions and the forces that shaped their evolution are largely unknown. There are two major difficulties in tracking them down. First, species that fail to produce any living or fertile hybrids cannot be studied through genetic crosses. And second, mate discrimination and hybrid sterility often result from the actions of many

genes, sometimes interacting in complex ways. So, even in species for which complete genome sequences are available, identifying the genes that cause speciation has proved difficult.

On page 715 of this issue, however, Presgraves and colleagues<sup>1</sup> describe how they mapped, identified and analysed one such gene. Found on chromosome 3 of the fruitfly *Drosophila simulans*, it interacts with one or more unknown genes on the *Drosophila melanogaster* X chromosome to cause the death of male hybrid offspring. In the two individual species, the protein products of these genes function quite happily together. But the gene on chromosome 3 has evolved divergently in the two species, such that it can no longer function with the X-chromosome genes of the other species in a hybrid male.

The authors began their study with fly strains lacking specific small segments of their genomes, and used them to map the hybrid-lethality effect to a single-copy gene



Figure 1 **Subjects of speciation research: *Drosophila melanogaster* and *Xiphophorus maculatus* (inset).**

on *D. simulans* chromosome 3. This gene encodes two nucleoporin proteins, Nup96 and Nup98, which are part of a large complex that is needed to shuttle RNAs and proteins between the nucleus and cytoplasm of cells.

Presgraves *et al.* rigorously confirmed their genetic mapping data by complementation testing: they performed individual crosses between *D. simulans* and strains of *D. melanogaster* (Fig. 1) that had mutations in each of 12 genes from the chromosomal region that includes the nucleoporin gene. Only those crosses involving *D. melanogaster* strains that had mutations in the nucleoporin gene failed to produce hybrid males, showing that this gene contributes to the death of hybrid males. Glazier *et al.*<sup>2</sup> have suggested that functional tests, such as complementation tests, are a prerequisite for identifying genes that contribute to complex traits, and Presgraves and colleagues' study satisfies this criterion. Presgraves *et al.* further showed that the *D. simulans* Nup96 protein — and probably one specific end of it, the amino-terminal region — bore the sequence variations that caused death when it was associated with the *D. melanogaster* X chromosome.

Two other genes that affect the fitness of hybrids have received much attention recently: the *Odysseus* gene in *Drosophila* species, and the *Xmrk-2* gene in *Xiphophorus* fish. *Odysseus* is thought to contribute to the sterility of hybrid males produced by crossing *D. simulans* with *D. mauritiana*<sup>3</sup>. It was mapped to a chromosomal region bearing a sequence known as a homeobox; genes containing this sequence are usually important in development. But five years after the identification of *Odysseus*, no functional tests have been

reported that support its role in hybrid sterility and its functional equivalence to the homeobox gene. So the particular homeobox gene that is purported to be *Odysseus* is still called *OdsH* (*Odysseus*-site homeobox gene), and Presgraves and colleagues' demonstration of the effect of the *Nup96* gene on hybrid fitness is more compelling.

The *Xmrk-2* gene, meanwhile, is over-expressed in hybrids produced by *Xiphophorus maculatus* platyfish and *X. helleri* swordtails, and causes invasive melanomas that can be fatal<sup>4</sup>. Functional tests have confirmed the role of this gene in tumour formation — but its effects are inconsistent when different strains of platyfish are used. It will be interesting to see whether *Nup96* from different strains of *D. simulans* also has variable effects on hybrid-male viability. The findings could help to determine the particular variations in the *Nup96* sequence that cause death.

One reason why evolutionary biologists seek specific genes associated with speciation is that, since Darwin's time, considerable debate has surrounded the exact role that natural selection plays in species formation. Statistical analyses of species-to-species variations in the sequence of a gene can indicate whether that gene was the target of natural selection: DNA sequences that are subject to recurrent 'positive' natural selection often exhibit more variations that alter the amino-acid sequence they encode than variations that do not. In today's era of genomics, many researchers have identified 'rapidly evolving' genes that appear to be targets of natural selection, and have suggested that these may be important in species formation. Until recently, however, there

has been little evidence that genes involved in speciation are necessarily direct targets of natural selection, or are evolving rapidly.

Presgraves *et al.*<sup>1</sup> also address this issue. The authors sequenced and analysed 15 copies of *Nup96* from different *D. simulans* and *D. melanogaster* individuals. Identifying differences in DNA sequences and putative amino-acid sequences both within and between species, they found the signature of positive natural selection early in the lineages of each species. The results show that the death of hybrid males produced by these species evolved at least in part through changes in the sequence of the *Nup96* protein, and not through changes in the sequence of genetic regulatory elements.

These findings accord with other data suggesting that the characteristics that define species are often the products of natural selection. Many analyses of genes involved in reproduction have identified the signature of natural selection<sup>5,6</sup>. For instance, sequence differences in the proteins encoded by some of these genes prevent fertilization between sperm and eggs of different species. Similarly, the *Drosophila Odysseus* gene sequences bear the signature of natural selection<sup>3</sup>. Finally, the *Hybrid male rescue* gene, which affects both hybrid viability and female sterility, also shows an extraordinarily large number of differences between *D. melanogaster* and *D. simulans*<sup>7</sup>. It seems from the emerging genetic data, then, that directional natural selection regularly affects genes that contribute to speciation.

This is an exciting time for speciation researchers. Studies such as that of Presgraves *et al.*<sup>1</sup> are finally identifying the precise genes that contribute to speciation and the forces that drove their evolution. A decade ago, the joke was that spell-checkers regularly attempted to substitute the word 'speciation' with 'speculation'. Now, the availability of numerous whole-genome sequences, the ability to survey patterns of expression for all genes in a genome simultaneously, and tremendous improvements in computational efficiency are bringing some of the most concrete advances in our understanding of the genetic changes that cause speciation. Speculation in this area will soon be a thing of the past. ■

Mohamed A. F. Noor is in the Department of Biological Sciences, Louisiana State University, Baton Rouge, Louisiana 70803, USA.

e-mail: mnoor@lsu.edu

1. Presgraves, D. C., Balagopal, L., Abmayr, S. M. & Orr, H. A. *Nature* **423**, 715–719 (2003).
2. Glazier, A. M., Nadeau, J. H. & Aitman, T. J. *Science* **298**, 2345–2349 (2002).
3. Ting, C.-T., Tsaur, S.-C., Wu, M.-L. & Wu, C.-I. *Science* **282**, 1501–1504 (1998).
4. Walter, R. B. & Kazianis, S. *ILAR J.* **42**, 299–321 (2001).
5. Swanson, W. J. & Vacquier, V. D. *Nature Rev. Genet.* **3**, 137–144 (2002).
6. Swanson, W. J. & Vacquier, V. D. *Annu. Rev. Ecol. Syst.* **33**, 161–179 (2002).
7. Barbash, D. A., Siino, D. E., Tarone, A. M. & Roote, J. *Proc. Natl Acad. Sci. USA* **100**, 5302–5307 (2003).

## Obituary

## Harmon Craig (1926–2003)

After the Second World War, weary of their part in the development of the atomic bomb, faculty at the University of Chicago turned to more esoteric matters. With mass spectrometers in place for making isotope measurements, Harold C. Urey and colleagues began studies in geochemistry and cosmochemistry. It was into this hotbed that Harmon Craig was propelled without waiting to finish his undergraduate degree. Craig, who died on 14 March, one day short of his seventy-seventh birthday, was to contribute hugely to these fields over the ensuing half-century.

One of the aims of the Urey group was to determine the temperatures of the ancient oceans. This depended on analysing carbon dioxide released from calcium carbonate fossils, and measuring the relative masses of the isotopes  $^{18}\text{O}$  and  $^{16}\text{O}$ . The constancy of the carbon isotopes was tacitly assumed. For his thesis, however, Craig measured the natural variability that occurs in the ratio of  $^{13}\text{C}/^{12}\text{C}$ . The independent discovery of natural radioactive  $^{14}\text{C}$  by W. F. Libby at the University of Chicago immediately led to dating applications in archaeology and geology. Craig's analysis permitted the proper determination of radiocarbon ages — later to be corrected to calendar years by accommodating the variations in initial  $^{14}\text{C}/^{12}\text{C}$ . His thesis remains the primary citation for work that involves  $^{13}\text{C}/^{12}\text{C}$  variation in natural materials — from studies of food chains to identifying sources of ancient marbles.

In cosmochemistry, the quest for the best measure of the early composition of the Solar System centres on meteorites called chondrites. It was commonly assumed that they have a uniform composition. Working with Urey, however, Craig first established 'quality certification' to reject meteorite samples modified by weathering, then showed that chondrites fall into at least two major groups. The Solar System, it emerged, was not so uniform after all, a discovery that with later work provided a new view of how and from what materials planets formed.

In 1955 came Craig's move to the institution where he remained for the rest of his career — the Scripps Institution of Oceanography in San Diego. The eastern universities in the United States were not yet ready to accept the strange world of geochemistry. In the west, however, the California Institute of Technology hired a bevy of geochemists from Chicago, while



## Pioneer in isotope geochemistry

Scripps, mainly through the foresight of Roger Revelle, its director, brought in Harmon Craig.

Back then, instruments were not built in a day. As Craig was tooling up, he tackled the question of what happens to  $\text{CO}_2$  in the atmosphere and the oceans. His theoretical solutions are valid to this day, and anticipated the programme for measuring atmospheric  $\text{CO}_2$  begun at Scripps by C. D. Keeling in 1957 at Revelle's instigation.

Craig decided that somebody should work out how oxygen and hydrogen isotopes are distributed in the hydrological cycle, especially if these isotopes were going to be used for palaeoenvironmental reconstructions. In two elegant papers that resulted from his presentation to an appreciative Italian audience at Spoleto in 1965, he laid out the framework for investigating kinetics and equilibrium in determining the isotopic composition of the hydrosphere, including the oceans. These 'Spoleto' papers remain fundamental documents in light-isotope geochemistry.

In 1967 came another turning point. At a meeting at the Woods Hole Oceanographic Institution, Henry Stommel proposed that a systematic study of the geochemistry and oceanography of all the oceans should be undertaken. With the new tracers and chronometers then available, it was the right moment to embark on this daunting enterprise. In due course it became clear that the leaders of this Geochemical Ocean Sections Study (GEOSECS), split among different institutions, should be Wallace Broecker, Derek Spencer and Craig. This

programme did not self-destruct, as some people thought (or maybe hoped) it would; it accomplished its main goals and spawned continuing projects. Craig himself was interested in the rate of turnover of the oceans. Here, the starting point was the idea that  $^{226}\text{Ra}$ , with a half-life of 1,620 years, might be a good tracer of circulation, being introduced at the ocean bottom from sediments and making its way up to the surface, decaying along the way. The  $^{226}\text{Ra}$  daughter,  $^{210}\text{Pb}$ , was proposed as a surrogate that could be measured. But when Craig and colleagues pursued this path, they discovered that  $^{210}\text{Pb}$  reacted with particles and was removed from the ocean by settling. This provided the basis for understanding the behaviour of particle-reactive elements in the ocean.

Another fruitful project concerned helium isotopes. The talented Brian Clarke developed a technique for measuring the ratio of  $^3\text{He}$  to  $^4\text{He}$ . When this was applied to an ocean water 'profile' collected by Craig, the astounding result was that, relative to atmospherically derived gases, there was an excess of  $^3\text{He}$ , not  $^4\text{He}$  as had been expected. The implication was that Earth's interior is still releasing a gas that was trapped when the planet was formed. The consequences lay not only in understanding ocean circulation but also in deciphering the way in which Earth's mantle behaves at ocean spreading centres and in ocean-island basalts.

In one of his last papers, Craig made sense of the cosmogenic  $^{32}\text{Si}$  measurements made on GEOSECS. To some, this was a hopelessly flawed set of data when tested with a simple model of the silica cycle. Craig and his co-authors responded with a paper, 'Paradox lost:  $^{32}\text{Si}$  and the global ocean silica cycle', in which the mixing of two sources of silica explained the results. So here was a man whose eye for high-quality measurements first showed up in studies on meteorites, but who decades later was instrumental in deciphering a major problem in marine geochemistry.

Harmon Craig influenced so many fields because he combined the strengths of a brilliant field observer with those of a meticulous measurer and a profound theorist. Yet this seeming one-man army was not acting alone. In everything he did he was aided and encouraged by his wife, Valerie. It was her patience with Craig's perennially searching mind that made the Craig enterprise such a success.

Karl K. Turekian

Karl K. Turekian is in the Department of Geology and Geophysics, Yale University, PO Box 208109, New Haven, Connecticut 06520-8109, USA. e-mail: karl.turekian@yale.edu



## Virology

### Pigs duck out of the influenza mix

*J. Virol.* **77**, 6988–6994 (2003)

In the twentieth century there were four worldwide influenza outbreaks, the most notorious being the 1918 pandemic, which killed an estimated 25 million people. Domesticated animals, notably poultry, are widely regarded as the source of that outbreak — and of any future ones. These animals carry a host of influenza viruses that, with relatively simple changes in their genetic sequence, can jump to humans. Genetic studies have suggested that another animal, the lead candidate being pigs, is needed as a ‘mixing vessel’ in which different bird flu viruses can recombine to produce a strain that infects humans.

But K. S. Li *et al.* suggest that pigs may not be necessary after all. Sampling the poultry markets of southern China, the authors found that one strain of bird flu, H9N2, can be passed back and forth between chickens and ducks. The strains seen in ducks had recombined twice and sometimes three times. And some ducks contained H9N2 viruses closely related to the H5N1 strain, which killed six people in Hong Kong in 1997 before it was eradicated by culling all of the territory’s chickens.

The authors warn that two-way viral transmission between domesticated ducks and chickens, which are often housed in close proximity, could produce influenza viruses with pandemic potential. **Tom Clarke**

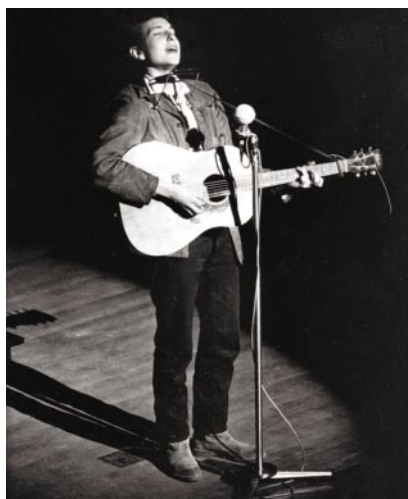
## Acoustics

### Guitar feedback tamed

*J. Acoust. Soc. Am.* **113**, 3188–3196 (2003)

Acoustic guitarists need no longer suffer from the feedback blues, thanks to a smart device developed by Steve Griffin and colleagues. The acoustic guitar is notoriously hard to amplify: its resonance tends to excite the unseemly howl of feedback. The problem can be mitigated by deadening the guitar body, but at the cost of ruining the sound of the unamplified instrument.

Griffin *et al.* have introduced active control of feedback by gluing two piezoelectric plates, just 0.25 mm thick, to the front plate of a guitar, below the bridge. The plates, which leave the unamplified tone untouched, are hooked up to the pick-up, the transducer that converts sound waves into an electrical signal for amplification. This control system, which could be powered by a small battery, kicks in only when the guitar is amplified. The piezoelectric plates vibrate, damping down the most problematic vibrational modes of the guitar body, as well as altering the wave shape of these



modes to reduce their coupling to the pick-up.

This active control system allows the guitar to be amplified to a level that is 7 decibels louder before feedback sets in. In a crowded pub, that could mean the difference between being heard and not. **Philip Ball**

## Gene regulation

### RNA switches

*Cell* **113**, 577–586 (2003)

Gene expression is commonly regulated by interactions between proteins and genes. Proteins can, for example, sense the levels of cellular metabolites and bind to the genes that regulate these levels, turning the production of the encoded enzymes on or off. But some metabolites can bind directly to messenger RNAs — the intermediate step between genes and enzymes. The RNA regions to which the metabolites bind are called riboswitches, and they too can control enzyme production.

Maumita Mandal and colleagues have now identified a new class of bacterial riboswitches that selectively recognize guanine, a purine base found in DNA and RNA. This type of riboswitch is encoded in at least five locations in the bacterial genome. These regions represent 17 genes that are involved in purine synthesis and transport. This and other research suggests that riboswitches contribute to gene regulation in bacteria, and help to regulate metabolism in living systems.

Some believe that riboswitches are derivatives of an ancient genetic control system that monitored metabolic and environmental signals long before proteins evolved. Guanine itself may be a relic of this primordial RNA world. So the identification of guanine as a riboswitch trigger is consistent with the authors’ theory that metabolite-sensing RNAs might have originated very early in evolution. **Helen R. Pilcher**

## Particle physics

### Brief lives

*Phys. Rev. Lett.* **90**, 203402 (2003)

The orthopositronium-lifetime puzzle is solved. So say R. S. Vallery and colleagues, who have measured how long this unstable particle survives, with unprecedented accuracy.

Orthopositronium is made of an electron and an anti-electron, or ‘positron’, bound together in matching spin states. Matter and antimatter annihilate each other, and so after a brief period of time — mere nanoseconds — the electron and positron disappear in a puff of  $\gamma$ -rays.

How long orthopositronium lives is predicted by quantum electrodynamics (QED), but previous measurements of its lifetime differed from the theoretical value by at least 0.1%. That might not sound like much. But QED is the most thoroughly tested theory in physics, and agreement between theory and experiment to more than a dozen decimal places is typical. In this context, 0.1% is not good enough.

To create orthopositronium, Vallery *et al.* fired a beam of positrons into a special nanoporous silica film, held in a vacuum cavity to protect against environmental influences. Their lifetime measurement is within 0.018% of the predicted value, and with more data should soon bring it to within 0.010%. **Alison Wright**

## Parasitology

### In the bag

*Proc. Natl Acad. Sci. USA*  
doi:10.1073/pnas.1131999100 (2003)

An insect parasite evades detection by cloaking itself in its host’s skin. The host is fooled into thinking that the parasite is part of its own body, leaving the intruder to devour its meal in peace.

Jeyarany Kathirithamby and colleagues have found this macabre camouflage in a group of insects called twisted-wing parasites, which belong to the order Strepsiptera. A strepsipteran larva punches a hole through its host’s tough outer skeleton. Then it pushes into the tissue layer beneath, creating a pocket that closes behind it to form a bag, inside which the larva feeds and develops. The grub is at first less than 0.1 mm long, but it grows to fill its host’s body. Females spend their entire lives here, pumping out new larvae. In mating, males fly around looking for the small part of a female’s body that protrudes from her host’s abdomen.

Cloaking themselves in host skin might be why the 596 species of Strepsiptera can infect a wide variety of hosts. For example, in one group, males target ants, but females develop inside grasshoppers. Most parasites are much more specialized. **John Whitfield**

# Super-tough carbon-nanotube fibres

These extraordinary composite fibres can be woven into electronic textiles.

The energy needed to rupture a fibre (its toughness) is five times higher for spider silk than for the same mass of steel wire, which has inspired efforts to produce spider silk commercially<sup>1–3</sup>. Here we spin 100-metre-long carbon-nanotube composite fibres that are tougher than any natural or synthetic organic fibre described so far, and use these to make fibre supercapacitors that are suitable for weaving into textiles.

We used a type of coagulation-based carbon-nanotube spinning method to prepare these fibres. In the original process<sup>4</sup>, surfactant-dispersed, single-walled nanotubes (SWNTs) were injected into a rotating bath of aqueous polyvinyl alcohol to produce nanotube gel fibres, which were then washed to desorb the polyvinyl alcohol and surfactant<sup>5,6</sup>. These fragile gel fibres were pulled from the coagulation bath at a rate of about 1 cm min<sup>-1</sup> in order to form solid nanotube fibres of some tens of centimetres in length<sup>6</sup>.

By modifying this method, we were able to spin a reel of nanotube gel fibre and then convert it into 100-m lengths of solid nanotube composite fibre in a continuous process, at a rate of more than 70 cm min<sup>-1</sup>. The spinning solution is injected into the centre of a cylindrical pipe in which the polyvinyl alcohol coagulation solution flows; contact with the coagulant collapses the spinning solution into a nanotube fibre associated with polyvinyl alcohol, which moves down the pipe for winding on a mandrel.

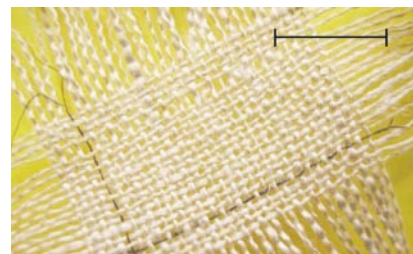
By using SWNTs synthesized from carbon monoxide<sup>7</sup> and lithium dodecyl sulphate as a surfactant, and by optimizing the flow rates of the spinning and coagulation solutions without removing polyvinyl alcohol from the gel, we are able to generate gel fibres that are mechanically strong enough for the second stage of the spinning

process. This stage involves unwinding the fibres onto a series of godets that carry them through an acetone-washing bath and then through a drying path so that they can be wrapped onto a mandrel.

The resulting composite fibres are about 50 µm in diameter and contain around 60% SWNTs by weight. They have a tensile strength of 1.8 gigapascals (GPa), which matches that of spider silk<sup>2</sup> (see supplementary information). This is seven times greater than the tensile strength of previously reported coagulation-spun nanotube fibres<sup>5</sup> and more than 10<sup>4</sup> times higher than that of 'dry-spun' carbon-nanotube fibres<sup>8</sup>. Our pre-drawn fibres match the energy absorption of spider silk up to the breaking strain of the toughest silk (30%), and continue absorbing energy until reaching an energy-to-break (570 J g<sup>-1</sup>) that is higher than that of spider dragline silk (165 J g<sup>-1</sup>; ref. 2, and see supplementary information), Kevlar fibre (33 J g<sup>-1</sup>; ref. 2) and graphite fibre (12 J g<sup>-1</sup>; ref. 9), and of previously described coagulation-spun nanotube fibres (1.7 J g<sup>-1</sup>; ref. 5). This toughness results from a combination of high strength and high strain to failure (Fig. 1).

Although, at 80 GPa, the Young's modulus of our nanotube composite fibres near failure strain is almost an order of magnitude less than that of high-performance graphite fibres or of individual SWNTs<sup>10,11</sup>, it is double that of previously described coagulation-spun SWNT fibres<sup>5</sup> and equal to that of directly synthesized 20-cm SWNT strands<sup>12</sup>, which have a strength of about 1.2 GPa and low toughness. Normalized for density, the Young's modulus and tensile strength of our fibres are more than twice the corresponding values for steel wire<sup>3</sup>, and our fibres are around 20 times as tough. This combination of high strength and high strain at break is not found for other reported nanotube composites<sup>13</sup>.

The elongation required for toughness is realized because of the absence of measurable fibre necking during deformation. This stability suggests that the strain rate should be roughly linearly dependent on the applied stress (newtonian flow), which we confirmed by stress-strain measurements (see supplementary information). Scanning electron micrographs reveal that the largely amorphous polyvinyl alcohol forms a coating on the nanotubes (results not shown), providing an important inter-phase region. The toughness of spider silk arises from chain extension in amorphous regions between relatively rigid crystalline protein blocks<sup>1</sup>, and amorphous polyvinyl



**Figure 2** Photograph of a textile containing two nanotube-fibre supercapacitors woven in orthogonal directions. Two nanotube composite fibres were separately coated with electrolyte by dipping in aqueous polyvinyl alcohol/phosphoric acid (19% phosphoric acid and 4% polyvinyl alcohol by weight), twisted together and then recoated with electrolyte. This fibre supercapacitor (diameter, 100 µm) provides a capacitance (5 F g<sup>-1</sup>) and energy-storage density (0.6 Wh kg<sup>-1</sup> at 1 V) that are comparable to those of large commercial supercapacitors; performance was unchanged over the measured 1,200 charge-discharge cycles. The helically wound nanotube fibres are separated at the capacitor ends so that electrical connections can be made. Scale bar, 1 cm.

alcohol between SWNTs may serve a similar function in our composite fibres. Slippage between individual nanotubes within bundles might also contribute to toughness.

We used our spun nanotube fibres to make supercapacitors and wove them into textiles (Fig. 2). Promising electronic-textile applications for these fibres, which are easy to weave and sew, include distributed sensors, electronic interconnects, electromagnetic shields, antennas and batteries.

**Alan B. Dalton\***, **Steve Collins\***, **Edgar Muñoz\***, **Joselito M. Razal\***, **Von Howard Ebron\***, **John P. Ferraris\***, **Jonathan N. Coleman†**, **Bog G. Kim\***, **Ray H. Baughman\***

*\*Department of Chemistry and The NanoTech Institute, University of Texas at Dallas Richardson, Texas 75080, USA*

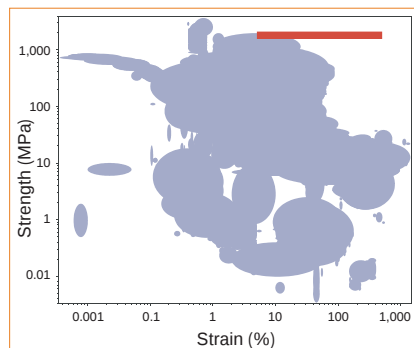
*e-mail: ray.baughman@utdallas.edu*

*†Department of Physics, Trinity College, Dublin 2, Ireland*

1. Kubik, S. *Angew. Chem. Int. Edn* **41**, 2721–2723 (2002).
2. Vollrath, F. & Knight, D. P. *Nature* **410**, 541–548 (2001).
3. Lesuer, D. R. *et al. Metallurg. Mater. Trans. A* **30**, 1559–1568 (1999).
4. Vigolo, B. *et al. Science* **290**, 1331–1334 (2000).
5. Vigolo, B., Poulin, P., Lucas, M., Luanos, P. & Bernier, P. *Appl. Phys. Lett.* **81**, 1210–1212 (2002).
6. Poulin, P., Vigolo, B. & Luanos, P. *Carbon* **40**, 1741–1749 (2002).
7. Nikolaev, P. *et al. Chem. Phys. Lett.* **313**, 91–97 (1999).
8. Jiang, K., Li, Q. & Fan, S. *Nature* **419**, 801 (2002).
9. Chand, S. J. *Mater. Sci.* **35**, 1303–1313 (2000).
10. Walters, D. A. *et al. Appl. Phys. Lett.* **74**, 3803–3805 (1999).
11. Yu, M.-F., Files, B. S., Arepalli, S. & Ruoff, R. S. *Phys. Rev. Lett.* **84**, 5552–5555 (2000).
12. Zhu, H. W. *et al. Science* **296**, 884–886 (2002).
13. Mamedov, A. A. *et al. Nature Mater.* **1**, 190–194 (2002).

**Supplementary information** accompanies this communication on Nature's website.

**Competing financial interests:** declared none.



**Figure 1** Comparison of the strength and failure strain for carbon-nanotube composite fibres for different degrees of initial pre-draw (red line) and the 3,000 materials of all types (lavender field) in the Cambridge Materials Selector database ([www.grantadesign.com](http://www.grantadesign.com)).

Avian metabolism

# Costs of migration in free-flying songbirds

**B**illions of songbirds migrate between continents twice each year, but the energy costs of this feat have never been measured for free-flying individuals. Here we follow New World *Catharus* thrushes during their nocturnal migratory flights<sup>1,2</sup>, recapturing individuals after journeys of up to 600 km and determining their energy expenditure by using doubly labelled water<sup>3,4</sup>. Although flight itself is costly, our results confirm the counterintuitive prediction that songbirds expend double the amount of energy during stopovers that they spend on flight over their entire migration<sup>5,6</sup>.

Quantifying the energy costs of migration should improve our understanding of songbird ecology<sup>7</sup>, but these costs have so far been estimated only from models<sup>8</sup> and wind-tunnel studies<sup>6,8,9</sup>. To measure migration costs for free-flying Swainson's (*Catharus ustulatus*) and hermit (*C. guttatus*) thrushes, we injected 38 birds (each weighing about 30 g; ref. 10) with 0.1 ml doubly labelled water (containing <sup>2</sup>H and <sup>18</sup>O isotopes) in the morning and followed 12 throughout one nocturnal spring-migration flight for up to 600 km or 7.7 h (Fig. 1a)<sup>12</sup>; we followed six of these birds until they landed at a stopover habitat. Injected birds that did not make a migratory flight that night were treated as controls (*n* = 26). The next morning, we recaptured birds at 9:10 (± 1 h 40 min) by mist-netting them and took a blood sample to quantify their energy expenditure<sup>4</sup>.

Our measurements indicate that both sustained natural flight and stopover are costly in cool weather (Fig. 1b, c). Migratory flight used 12.5 kJ h<sup>-1</sup>, determined as flying cost above resting metabolism, or 15.5 kJ h<sup>-1</sup> (4.3 watts) total energy while flying, which agrees with predicted values<sup>5,6,8</sup>, the cost of transport is 1.1 (calculated as a flight power of 4.3 watts divided by a flight speed of 13 m s<sup>-1</sup> and a body weight of 0.3 N (equivalent to a body mass of 30 g)), which results in a mechanical cost of transport of about 0.25 (1.1 × 0.23, where 0.23 is the energy-conversion efficiency).

These results for free-flying birds validate a generation of theoretical models<sup>5,8</sup> and wind-tunnel studies<sup>6,11</sup>. Surprisingly, stopover thrushes that did not fly on cold nights incurred increased thermoregulation costs comparable to those of a 2.5-h flight (Fig. 1b), illustrating the importance of climate during spring migration (Fig. 1c). Three of six migrating birds chose to stop over within 50 m of houses in urban or suburban areas.

We found that individual birds had roughly the same body weight and fat content in the mornings before and after their migratory flights (6% body-weight loss, no change

in fat content), indicating that they probably regulate the duration of a migratory flight according to their body weight and fat stores<sup>6</sup>.

We can now estimate the energy expenditure of spring migration in a songbird<sup>11</sup>. *Catharus* thrushes can travel from Panama to Canada in 42 days<sup>10</sup>. Only 18 of 42 days include nocturnal flights, which are 265 km on average (4.6 h in flight; Fig. 1b). The average cost for migration days is 130 kJ, of which 71.3 kJ is spent on flight alone; the remaining 24 days are stopovers where an average of 88 kJ d<sup>-1</sup> is expended, assuming an ambient temperature similar to that in central Illinois. For a 4,800-km migration, about 4,450 kJ is spent — an overall cost of some 0.93 kJ km<sup>-1</sup>.

Our measurements indicate that actual flight represents only 29% of total energy expenditure during the entire migration, confirming the predicted 1:2 ratio of energy costs during flight versus stopover<sup>6</sup>. These calculations do not take into account adverse weather conditions, which could strongly affect the energy balance, or the wide variation in stopover costs for different habitats<sup>7</sup>. Days that include migratory flight (130 kJ d<sup>-1</sup>) are significantly less expensive than days spent feeding large chicks, for example<sup>12</sup>.

We suggest that the timing of migration in relation to climate, particularly the ambient temperature, is important for birds that use twice as much energy during stopover as during migratory flight. Springtime reverse migration<sup>13</sup> can now be better understood as an energy-saving strategy to minimize thermoregulation costs.

**Martin Wikelski**\*†‡, **Elisa M. Tarlow**†, **Arlo Raim**‡, **Robert H. Diehl**†, **Ronald P. Larkin**‡, **G. Henk Visser**§¶

\*Department of Ecology and Evolutionary Biology, Princeton University, New Jersey 08544, USA

e-mail: wikelski@princeton.edu

†Department of Animal Biology, University of Illinois at Urbana-Champaign, Illinois 61801, USA

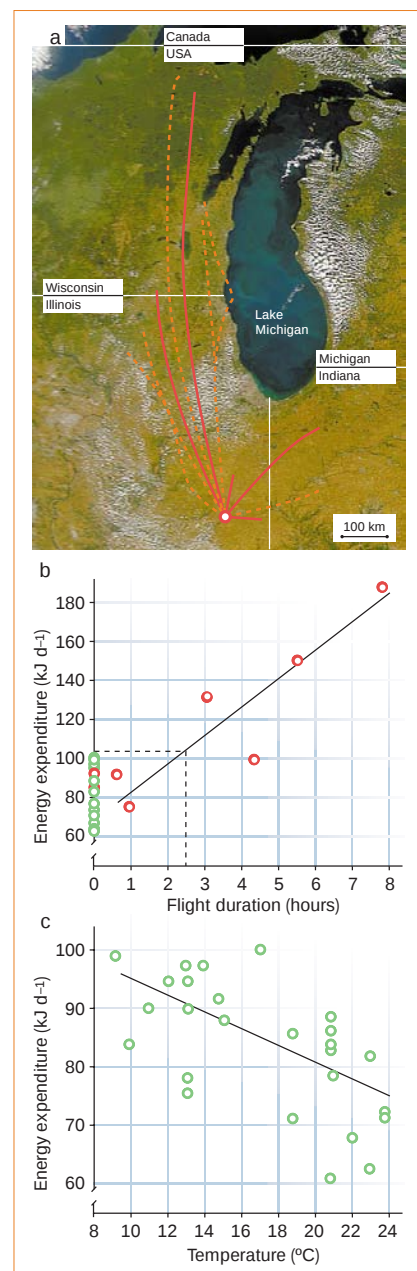
‡Illinois Natural History Survey, Champaign, Illinois, USA

§Centre for Isotope Research, Groningen University, 9747 AG Groningen, The Netherlands, and

¶Zoological Laboratory, Haren, The Netherlands

1. Cochran, W. in *Animal Orientation and Navigation* (eds Galler, S. R., Schmidt-Koenig, K., Jacobs, G. J. & Belleville, R. E.) 39–59 (NASA, Washington DC, 1972).
2. Cochran, W. *Anim. Behav.* **35**, 927–929 (1987).
3. Nagy, K. A. *Am. J. Physiol.* **238**, 454–473 (1980).
4. Speakman, J. R. *Doubly Labeled Water: Theory and Practice* (Chapman and Hall, London, 1997).
5. Hedenström, A. & Ålerstam, T. *J. Theor. Biol.* **189**, 227–234 (1997).
6. Ålerstam, T. & Hedenström, A. *J. Avian Biol.* **29**, 343–369 (1998).
7. Faaborg, J., Brittingham, M., Donovan, T. & Blake, J. in *Ecology and Management of Neotropical Migratory Birds: A Synthesis and Review of Critical Issues* (eds Martin, T. E. & Finch, D. M.) 357–380 (Oxford Univ. Press, New York, 1995).
8. Pennycuik, C. J. *J. Theor. Biol.* **191**, 47–61 (1998).
9. Klaassen, M. *J. Exp. Biol.* **199**, 57–64 (1996).
10. Graber, R. R., Graber, J. W. & Kirk, E. L. *Biol. Notes No. 75* (Illinois Natural History Survey, Champaign, Illinois, 1971).
11. Masman, D. & Klaassen, M. *Auk* **104**, 603–616 (1987).
12. Bennett, P. & Harvey, P. H. *J. Zool.* **213**, 327–363 (1987).
13. Richardson, W. J. *Oikos* **30**, 224–272 (1978).

**Competing financial interests:** declared none.



**Figure 1** Migration routes and energetics of Swainson's and hermit thrushes during April and May in 1999 and 2000 (refs 1, 2, 10). We pooled data for final analysis because all traits were statistically indistinguishable between species<sup>10</sup>. **a**, Tracks for 12 individuals starting from Urbana, Illinois: full lines, recaptured birds; dashed lines, lost birds. Ends of lines indicate stopover sites. **b**, Daily energy expenditure (DEE) of migrating (red symbols) and stopover (green symbols) birds; for migrating birds, this is described as a linear regression:  $DEE (kJ) = 70 (\pm 9) + 12.5 (\pm 2.1) \times \text{flight duration in hours}$ ;  $r^2 = 0.89$ ,  $P = 0.004$  (standard deviations are given in parentheses) (ref. 11). Dashed line, average flight duration below which energy expenditure is indistinguishable from stopover controls (2.5 h). **c**, Variability in energy expenditure of birds during stopover is explained by differences in average daily ambient temperature ( $DEE (kJ) = 109.6 (\pm 6.5) - 1.52 (\pm 0.37) \times \text{ambient temperature}$ ;  $r^2 = 0.4$ ,  $n = 26$ ,  $P < 0.001$ ). We calibrated measurements using doubly labelled water with five birds held in a respirometer for 24 h (accuracy of ± 4% according to equation 7.17 of ref. 4).



# Reticular synthesis and the design of new materials

Omar M. Yaghi\*<sup>†</sup>, Michael O'Keeffe\*<sup>‡</sup>, Nathan W. Ockwig\*<sup>†</sup>, Hee K. Chae\*<sup>†§</sup>, Mohamed Eddaoudi\*<sup>†</sup> & Jaheon Kim\*<sup>†</sup>

\* Materials Design and Discovery Group, <sup>†</sup> Department of Chemistry, 930 N. University Avenue, University of Michigan, Ann Arbor, Michigan 48109-1055, USA

<sup>‡</sup> Department of Chemistry, Arizona State University, Tempe, Arizona 85287-1604, USA

<sup>§</sup> Permanent address: Hankuk University of Foreign Studies, Korea

**The long-standing challenge of designing and constructing new crystalline solid-state materials from molecular building blocks is just beginning to be addressed with success. A conceptual approach that requires the use of secondary building units to direct the assembly of ordered frameworks epitomizes this process: we call this approach reticular synthesis. This chemistry has yielded materials designed to have predetermined structures, compositions and properties. In particular, highly porous frameworks held together by strong metal–oxygen–carbon bonds and with exceptionally large surface area and capacity for gas storage have been prepared and their pore metrics systematically varied and functionalized.**

**A**lthough the synthesis of new materials has long been recognized as the most essential element in advancing technology, it generally remains more of an art than a science—in that the discovery of new compounds has mostly been serendipitous, using methods referred to by critics as ‘shake and bake’, ‘mix and wait’ and ‘heat and beat’. For much of the twentieth century, this worked well for the synthesis of important solid-state materials, and we expect that it will continue to yield interesting compounds. However, it is becoming increasingly urgent to produce materials designed to perform highly specific and cooperative functions<sup>1,2</sup>.

Recent extensive research into the design and synthesis of metal-organic frameworks (MOFs) has led to numerous practical and conceptual developments in that direction<sup>3–7</sup>. Specifically, the chemistry of MOFs has provided an extensive class of crystalline materials with high stability, tunable metrics, organic functionality, and porosity. Here we present some of the important developments that have shaped this rapidly growing field and propose a general conceptual framework, which serves as a useful tool in designing materials constructed from molecular building blocks.

## Reticular synthesis

General lack of control over the character of solids produced from the traditional synthetic methods is directly related to the fact that the starting entities do not maintain their structure during the reaction, leading to poor correlation between reactants and products. We have shown that the design of an extended network can be realized by starting with well-defined and rigid molecular building blocks that will maintain their structural integrity throughout the construction process<sup>3,8</sup>. Alternatively, the use of well-defined conditions that lead to the formation of such building blocks *in situ* is an equally viable approach to the design of extended structures and one that has been especially fruitful in MOF chemistry<sup>8</sup>. We have shown<sup>8,9</sup> that the practice of logical synthesis must begin with knowledge of the target network ‘blueprint’ and identification of the required building blocks for its assembly. This process is central to our ability to achieve true design of solid-state materials: we refer to its implementation as reticular synthesis.

In essence, reticular synthesis can be described as the process of assembling judiciously designed rigid molecular building blocks into predetermined ordered structures (networks), which are held together by strong bonding. It is different from retrosynthesis of organic compounds<sup>10</sup>, because the structural integrity and rigidity of the building blocks in reticular synthesis remain unaltered

throughout the construction process—an important aspect that could help to realize fully the benefits of design in crystalline solid-state frameworks. Similarly, reticular synthesis should be distinguished from supramolecular assembly<sup>11</sup>, because in the former, building blocks are linked by strong bonds throughout the crystal.

In this context, it is worth noting that a plethora of extended structures successfully prepared by copolymerization of metal ions with organic links have been reported and reviewed under the rubric of ‘crystal engineering’, a term which has mutated somewhat over the years and which is now rather broadly defined<sup>12–14</sup>. We use the more precise term reticular synthesis (or chemistry) to describe the logical approach to the synthesis of robust materials with pre-designed building blocks, extended structures, and properties. Thus, it might be considered a subclass of crystal engineering.

Research in the area of MOFs and, to a lesser extent, inorganic frameworks, has recently progressed far enough to provide examples which show the feasibility of the building-block approach and the scope of reticular synthesis. Here we take selected examples to highlight the ideas involved, with emphasis on our own work on MOFs in which structural and thermal stability characterization have been combined with in-depth studies of surface area and gas sorption capacities. The subject of metal-organic frameworks has a long history, however, so we begin by giving some perspectives on the field.

## Perspectives and challenges

The Cambridge Structure Database (CSD) documents the crystal structures of more than 11,000 extended metal-organic compounds in which a metal ion or cluster has been linked by an organic moiety in which the linking functionality is a cyanide, pyridyl, phosphate or carboxylate. Of these, nearly 3,000 compounds have three-dimensional (3D) structures and about double that number have 2D structures. MOFs composed of the first-row transition metals and organic links such as cyanide<sup>15,16</sup>, glutamate<sup>17</sup>, formate<sup>18</sup>, triazole<sup>19</sup>, oxalate<sup>20</sup>, 1,2,4,5-tetracarboxylates<sup>21</sup> and squarates<sup>22</sup> have long been known. Also, expanded diamond structures based on linking of Cu(I) with ANT<sup>23</sup>, DCNQ<sup>24</sup> and TCTPM<sup>25</sup>, and decorated diamond structures based on linking Mn(II) with Ge<sub>4</sub>S<sub>10</sub><sup>4–</sup> (T2) clusters<sup>9</sup> (see Box 1 for abbreviations of chemical names and clarification of terms) represent the numerous forerunners of the extended structures constructed from molecular building blocks. Subsequently, other compounds based on linking metal ions with ADC<sup>26</sup>, ETC<sup>27</sup>, INT<sup>28</sup>, BPY<sup>29,30</sup> and BPCN<sup>31</sup> have been reported. Tetrahedral building blocks such as HMT<sup>32</sup> and

MCP<sup>33</sup> have been linked by metal ions to form decorated diamond structures, while TPY<sup>34</sup> and ATC<sup>35</sup> are examples of links pre-designed so that their assembly into diamond networks is done by linking through hydrogen bonding.

Consideration of this range of crystal structures, in particular those of MOFs, has led us to identify two critical questions for reticular synthesis. First, of the almost unlimited possible networks, which can be expected to form and how can they be synthesized?

## Box 1

### Definitions and structures

Here we define some terms and illustrate organic species referred to in the text by abbreviations in the Box 1 figure.

**Reticular:** (adjective) having the form of a (usually periodic) net.

**Isorecticular:** Based on the same net (having the same topology).

**MOF-*n*:** Metal-organic framework (with *n* an integer assigned in roughly chronological order).

**IRMOF-*n*:** Isorecticular MOF (with *n* an integer referring to a member of the series).

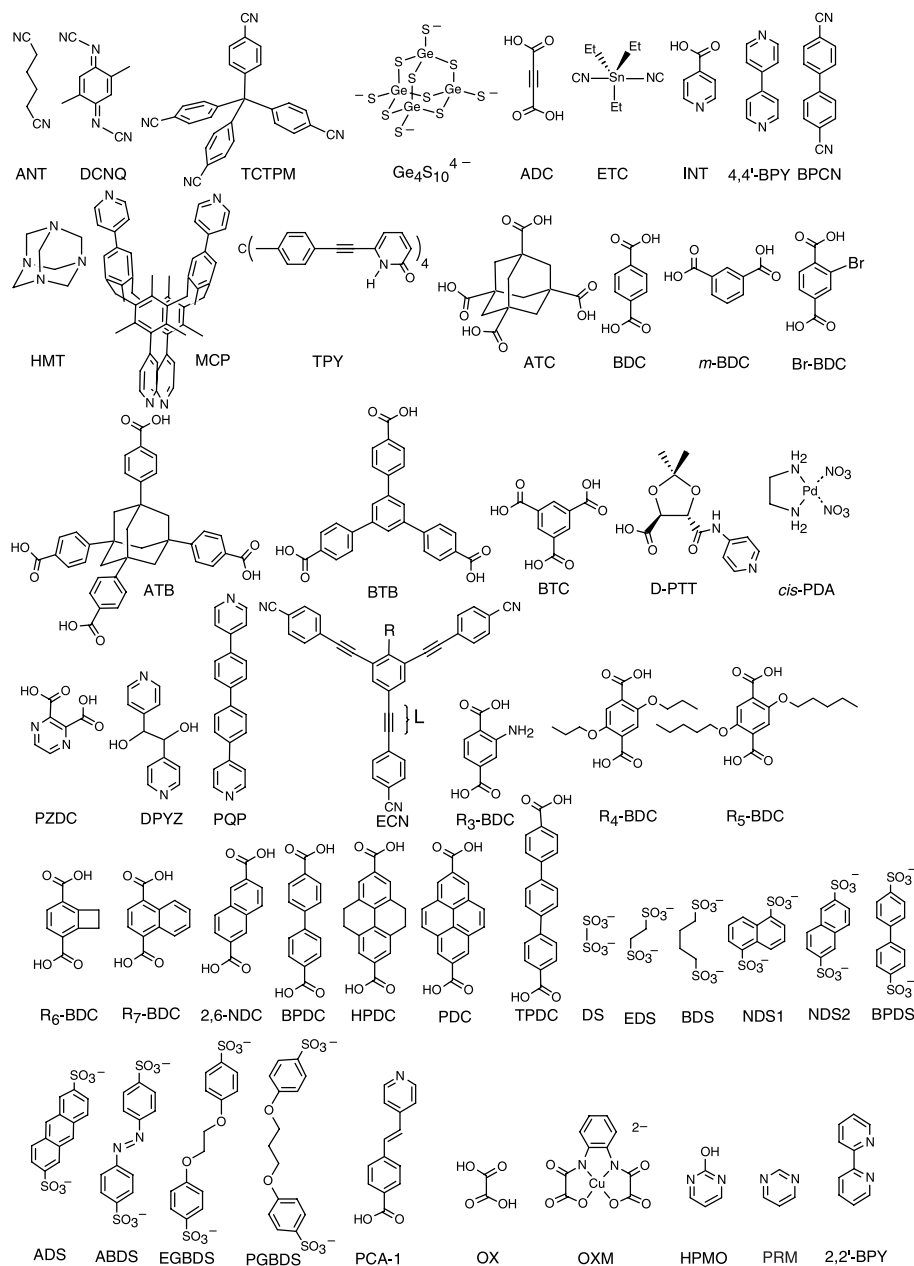
**Interpenetration:** A term used to describe the mutual intergrowth of two or more networks in a structure where the networks are physically

but not chemically linked. See refs 4 and 8 for reviews on this property.

**Expansion:** Increasing the spacing between vertices in a network.

**Decoration:** Replacing a vertex in a net by a group of vertices.

**SBU:** The term 'secondary building unit' has been used for some time to describe conceptual fragments of zeolites; in the context of reticular chemistry it refers to the geometry of the units defined by the points of extension (such as the carboxylate C atoms in most carboxylate MOFs).



Second, with few exceptions, MOFs based on M-N linkages in which the vertex of the network is just a single atom have a tendency to form structures which collapse upon removal of 'guest' atoms from the pores, rendering the structure nonporous. How then can we ensure that our structures will be robust?

### Secondary building units

To address these questions it is important to first appreciate that it is difficult (although not impossible) to attempt *a priori* synthesis of structures, such as those presented above, from simple metal ions and organic links because ions hold little directional information. This relative lack of directionality often results in flexibility around the metal ion, a multiplicity of possible structures and a general lack of control—as exemplified by frameworks based on Cu ions and bipyridine and related links<sup>13,36–42</sup>. We have focused on using the carboxylate functionality to chelate metal ions and lock them into rigid and thus directional metal–oxygen–carbon clusters, with the

points of extension (in this example, carboxylate C atoms) defining geometrical shapes referred to as secondary building units (SBUs)<sup>26</sup>. The successful design of rigid frameworks based on such SBUs was demonstrated for the first time in MOF-2 (see ref. 43 and below) and MOF-5<sup>44</sup>. The design principles employed for MOF-5 will be discussed here to show that SBUs have intrinsic geometric properties that facilitate network design and help us to address the issues of network synthesis and robustness.

In MOF-5 (Fig. 1a),  $\text{Zn}_4\text{O}(\text{CO}_2)_6$  units containing four  $\text{ZnO}_4$  tetrahedra with a common vertex and six carboxylate C atoms that define an octahedral SBU are joined together by benzene links. This leads to a cubic network in which the vertices are the octahedral SBUs and the edges are the benzene struts. In practice, this compound was prepared from Zn(II) and BDC acid under conditions pre-determined to yield the octahedral SBU *in situ*. Because the SBU and benzene links are relatively large and rigid entities, the structure produced has exceptional porosity (indicated by its sorption), and stability (indicated by thermal analysis and single crystal X-ray diffraction studies on the completely evacuated framework<sup>8,44</sup>).

The exceptional stability of MOF-5 can be understood by comparing its basic network, composed of single atom vertices (Fig. 1b), with the actual structure of MOF-5, which has cationic clusters at those vertices (Fig. 1a). The basic network has no resistance to shear if the links are considered to be universal joints. However, in the actual MOF-5 structure, the cationic clusters have a truncated tetrahedral envelope (Fig. 1c), and the rigidly planar  $\text{O}_2\text{C}-\text{C}_6\text{H}_4-\text{CO}_2$  linkers have a planar slat envelope. The linkage of these two groups produces an inherently rigid structure held together by mutually perpendicular hinges.

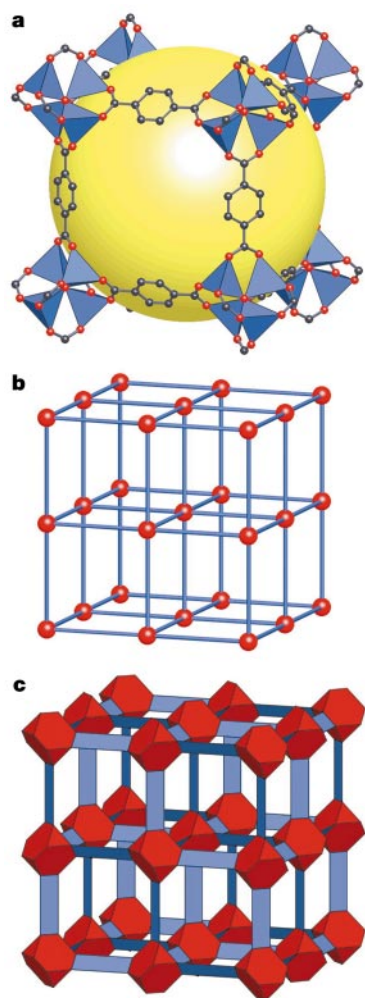
This approach, based on the concept of SBUs, has been useful in rationalizing the topologies of MOF structures<sup>26</sup>, and more importantly, it has allowed the synthesis and use of a large number of inorganic and organic SBUs with varying geometries (Fig. 2)<sup>26,43–52</sup>. In many of these cases, identifying the reaction conditions that produce an SBU with a specific geometry *in situ* means that the addition of a rigid organic SBU will result in the formation of a pre-determined network. In other words, with this strategy it is now possible to control the overall coordination number of the inorganic and organic SBUs, and therefore the need to identify the networks that are expected to form from different geometric shapes becomes particularly acute.

### Possible framework topologies

Which framework topology can be expected to form is of critical importance because there are a large number of possible structures for each geometrical shape. For example, more than 100 different topologies (in addition to that of diamond)<sup>53</sup> are possible for linking tetrahedral building blocks together into structures with just one kind of vertex (that is, all vertices related by symmetry), and the would-be designer must find some criteria for deciding which will be selected.

We have recently argued that in general only a small number of simple, high-symmetry structures will be of overriding general importance, and they would be expected to form most commonly<sup>54</sup>. We call these default structures, in that they would be expected to form if the assembly reaction involved simple building blocks, especially single atoms (ions) joined by flexible linkers. More complicated structures can be targeted by judicious use of appropriately shaped SBUs and linkers (we give examples below). A list of some default structures<sup>54</sup> and some criteria for their classification<sup>55–60</sup> are given in Box 2.

We have undertaken a systematic survey<sup>59</sup> of structures with the goal of design in mind. There are just six three-periodic structures with one kind of vertex, edge and ring—we call these regular and quasi-regular. The next most regular are those with just one kind of link. Their enumeration is mostly straightforward and—together



**Figure 1** The MOF-5 structure and its topology. **a**, The MOF-5 structure shown as  $\text{ZnO}_4$  tetrahedra (blue polyhedra) joined by benzene dicarboxylate linkers (O, red and C, black) to give an extended 3D cubic framework with interconnected pores of 8 Å aperture width and 12 Å pore (yellow sphere) diameter. (Yellow sphere represents the largest sphere that can occupy the pores without coming within the van der Waals size of the framework). **b**, The topology of the structure (primitive cubic net) shown as a ball-and-stick model. **c**, The structure shown as the envelopes of the  $(\text{OZn}_4)\text{O}_{12}$  cluster (red truncated tetrahedron) and benzene dicarboxylate (BDC) ion (blue slat). Note that opposing slats are all at 90°.

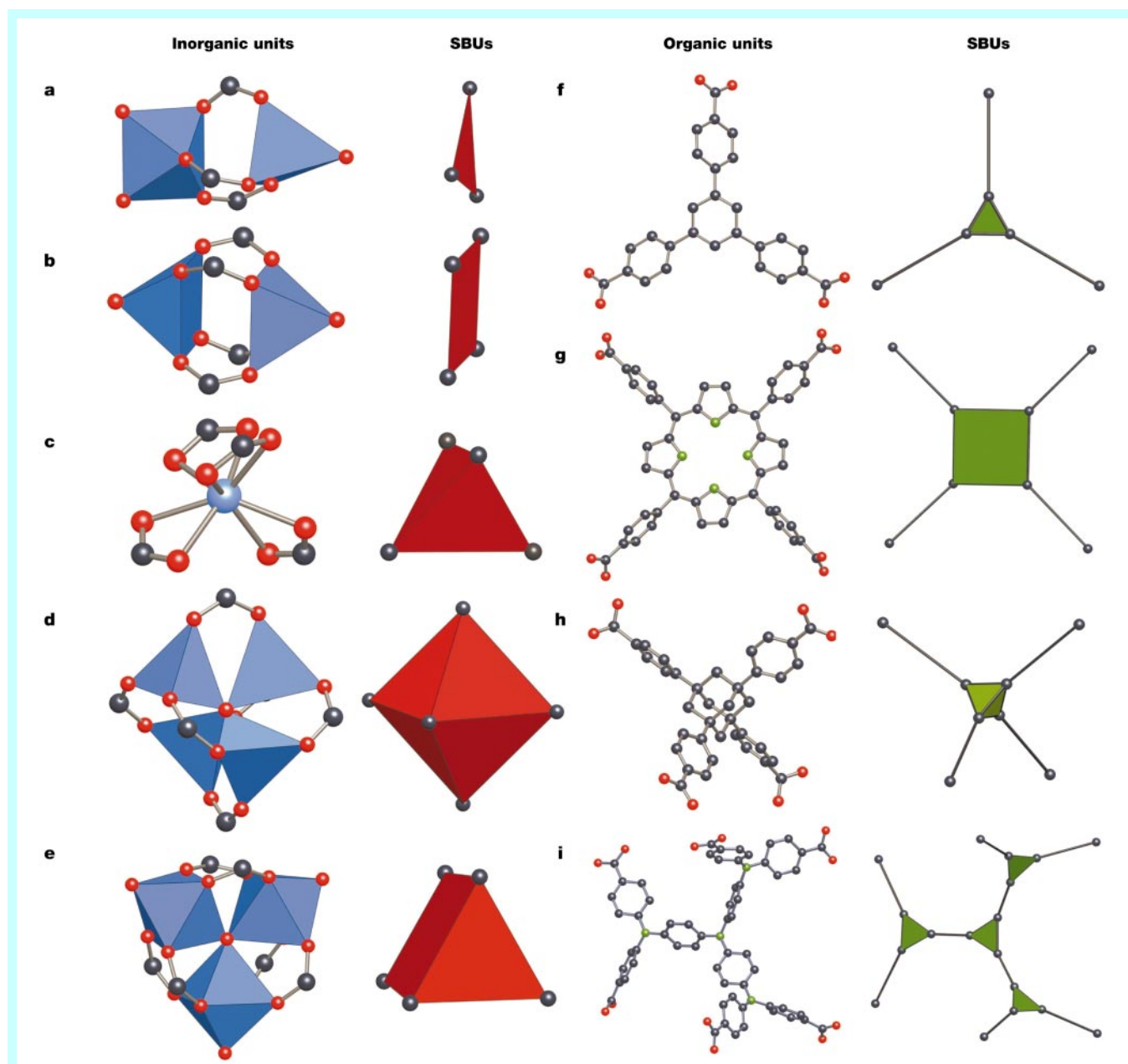


with the regular structures—they are the most common structures used in reticular chemistry, because one almost always attempts a synthesis with just one kind of linker. As an illustration, we<sup>48</sup> have enumerated the nine principal ways in which square SBUs can be joined by equal linkers to form polyhedra, rods, layers or three-dimensional nets. In the next section we show how this enumeration allows particular linked-square structures of each different dimensionality to be successfully targeted for synthesis using appropriately shaped linkers.

### Design and synthesis of pre-determined frameworks

As we noted above, metal-containing (cationic) SBUs are not isolatable entities, and so it has been essential to establish the exact chemical conditions that will yield a specific SBU *in situ*. On the other hand, organic (anionic) SBUs are pre-assembled

using the considerable sophistication of organic synthetic methods<sup>10</sup>—see Box 1 and Fig. 2 for examples. For a given cationic SBU the geometry of the organic unit plays an important role in directing the structure. This is illustrated (Fig. 3) for four structures<sup>41,61–63</sup> of different dimensionality (periodicity—0-D refers to an object within only point symmetries): a truncated cuboctahedron (0D), a linear rod (1D) the square grid (2D), and the NbO network (3D) based on linkage of paddle-wheel clusters by ditopic linkers. The dihedral angle between the paddle wheels (square SBUs) plays a crucial role in determining which one of these structures will be produced. The reaction conditions that would yield the paddle-wheel clusters were used in the presence of the appropriate link, to give the desired structure—which is in each case the only possible structure given identical linkers of the shape indicated.



**Figure 2** Examples of SBUs from carboxylate MOFs. O, red; N, green; C, black. In inorganic units<sup>9</sup> metal-oxygen polyhedra are blue, and the polygon or polyhedron defined by carboxylate carbon atoms (SBUs) are red. In organic SBUs the polygons or polyhedra to which linkers (all  $-\text{C}_6\text{H}_4-$  units in these examples) are attached are

shown in green. The geometry for **i** is an example of a tertiary building unit<sup>52</sup> consisting of four SBUs (green triangles). The carboxylate carbons of this unit are at the vertices of a trigonal prism.

This approach can be used in linking together other SBU shapes. For example, the PtS net is the default structure for linking squares (Pt) and tetrahedra (S). Assembly of the paddle wheel with ATC and ATB (tetrahedral SBUs) yields MOF-11<sup>51</sup> and MOF-36<sup>26</sup>, which are indeed based on the PtS net. There are only two ways of linking square and triangular SBUs—these are the nets named for Pt<sub>3</sub>O<sub>4</sub> and ‘twisted boracite’—and they might both be considered default structures<sup>54</sup> (Box 2). Again, linking a copper-based paddle wheel with triangular SBUs, either BTB or BTC, gives MOF-14<sup>49</sup> with the Pt<sub>3</sub>O<sub>4</sub> topology and HKUST-1<sup>64</sup> with the twisted boracite topology, respectively. Examination of the parent nets shows that the simpler Pt<sub>3</sub>O<sub>4</sub> net requires the more flexible BTB links.

The regular (and hence default) structure for linking triangles is the chiral 3D Si network in SrSi<sub>2</sub>, which has been a target of reticular synthesis<sup>45,65</sup>. Interestingly, it is composed of achiral triangular vertices, which twist with respect to adjacent like vertices by a dihedral angle of 70.5° to make an overall chiral arrangement. This topology was assembled from BTC and Zn(II) under conditions

that give a propeller-shaped binuclear M<sub>2</sub>(CO<sub>2</sub>)<sub>3</sub> SBU (Fig. 2a)<sup>45</sup>. Here, three carboxylate C atoms of the SBU are twisted with respect to the C atoms on the benzene ring to give chiral crystals although the bulk material is racemic. Recently, it has been demonstrated that an enantiopure MOF can be prepared from a derivative *D*-tartaric acid (D-PTT) and Zn(II) to give a triangular SBU, which formed a hexagonal layered network<sup>66</sup> that was functionalized with pyridine units to allow the pores to be used for chiral catalysis.

It is worth noting that exquisite control over the outcome of reticular synthesis has also been developed using site-blocked metal complexes that express angular information. For example, *cis*-PDA provides 90° binding angles such that with linear ditopic organic links such as BPY<sup>67</sup>, the nitrate ligands are displaced by BPY, leading to molecular squares. On the basis of this approach other metal complexes of various binding angles and a variety of branching and bent organic links have been used in the design of other polygons, platonic and archimedean polyhedra<sup>61,68–71</sup>.

## Box 2

### Framework topologies

The topology underlying periodic structures is that of a periodic net. Crystal chemists, particularly Wells<sup>15</sup>, have laboured to enumerate, describe and classify such nets. We<sup>58</sup> find the approach based on tiling to be particularly fruitful<sup>55,66</sup>. The tiles are generalized polyhedra (cages) which generate the entire structure when packed together. For each periodic net there is a unique ‘natural’ tiling<sup>58</sup>; examples are shown in Fig. 5. The adamantane unit with ten vertices is the natural tile of the diamond net. Its four faces are six-membered rings, which we symbolize as [6<sup>4</sup>]. The natural tiling of the lonsdaleite net has equal numbers of two tiles that are [6<sup>3</sup>] and [6<sup>5</sup>] (Fig. 5).

To classify nets we make use of the transitivity of the structure<sup>57</sup>. If there are *p* kinds of vertex, *q* kinds of edge (link), *r* kinds of ring (faces of tiles) and *s* kinds of tile, the transitivity is *pqrs*. It turns out<sup>58</sup> that the five structures whose natural tilings have transitivity 1111 have coordination figures that are regular polygons or polyhedra and we call them ‘regular’ (see Table 1). Their importance in reticular synthesis is comparable to that of the five platonic solids in other areas of chemistry. Structures 1112 are called ‘quasiregular’ and structures 11rs are ‘semiregular’. These are the structures that are most likely to form with one kind of SBU joined by one kind of link. They can be enumerated fairly easily and there are only about twenty

of them (the exact number depends on some subtle details of what we consider to be allowable structures). There are another twenty or so structures with two vertices and one edge. Structures with just one kind of edge (edge transitive) are the most reasonable targets of designed reticular synthesis, and thus the most important from that point of view. Even these approximately 50 structures can quickly be reduced to a smaller number. For example, there are seven edge-transitive nets with tetrahedral vertices (examples are given in Table 1) but the diamond net is the most regular, and thus the default net for tetrahedral coordination.

Of course less regular nets can be forced by low-symmetry SBUs. For example, 5-coordinated SBUs cannot be linked by just one kind (that is, symmetry equivalent) of link. Two simple examples are given in Table 1. There is only one 3-coordinated 3D net with symmetry that requires the coordination figure to be an equilateral triangle (that of SrSi<sub>2</sub>—see Table 1). The use of 3-coordinated SBUs with a T or Y shape can accordingly lead to a lower-symmetry net such as that of Si in ThSi<sub>2</sub> (also listed). The formation of the rutile net instead of the more symmetrical pyrite net in (6,3)-coordinated structures is often ascribable to similar causes.

Table 1 Some important nets and their natural tilings

| CN  | LC  | Name                           | Vertex figure         | Transitivity | Tiles                                                                  |
|-----|-----|--------------------------------|-----------------------|--------------|------------------------------------------------------------------------|
| 3   | +Y* | SrSi <sub>2</sub>              | Triangle              | 1111         | [10 <sup>3</sup> ]                                                     |
| 4   | J*  | NbO                            | Square                | 1111         | [6 <sup>8</sup> ]                                                      |
| 4   | D   | Diamond                        | Tetrahedron           | 1111         | [6 <sup>4</sup> ]                                                      |
| 6   | cP  | Primitive cubic                | Octahedron            | 1111         | [4 <sup>6</sup> ]                                                      |
| 8   | cl  | Body-centred cubic             | Cube                  | 1111         | [4 <sup>4</sup> ]                                                      |
| 12  | cF  | Face-centred cubic             | Cuboctahedron         | 1112         | 2[3 <sup>4</sup> ] + [3 <sup>8</sup> ]                                 |
| 3   |     | ThSi <sub>2</sub>              | T or Y                | 1211         | [10 <sup>4</sup> ]                                                     |
| 4   |     | CdSO <sub>4</sub>              | Square                | 1221         | [6 <sup>2</sup> .8 <sup>2</sup> ]                                      |
| 4   | W*  | Sodalite                       | Tetrahedron           | 1121         | [4 <sup>8</sup> .6 <sup>8</sup> ]                                      |
| 4   | Q   | Quartz                         | Tetrahedron           | 1121         | [6 <sup>2</sup> .8 <sup>2</sup> ]                                      |
| 4   |     | Lonsdaleite                    | Tetrahedron           | 1222         | [6 <sup>3</sup> ] + [6 <sup>5</sup> ]                                  |
| 6   | E   |                                | Trigonal prism        | 1122         | 2[4 <sup>3</sup> ] + [4 <sup>3</sup> .6 <sup>2</sup> ]                 |
| 5   |     | CaB <sub>6</sub>               | Square pyramid        | 1222         | [3 <sup>7</sup> ] + [3 <sup>3</sup> .4 <sup>9</sup> ]                  |
| 5   |     | BN                             | Trigonal bipyramid    | 1221         | [4 <sup>6</sup> .6 <sup>2</sup> ]                                      |
| 6   |     | Polybenzene                    | Hexagon               | 1121         | [4 <sup>6</sup> .6 <sup>4</sup> ]                                      |
| 4,8 |     | Fluorite (CaF <sub>2</sub> )   | Tetrahedron, cube     | 2111         | [4 <sup>12</sup> ]                                                     |
| 3,6 |     | Pyrite (FeS <sub>2</sub> )     | Triangle, octahedron  | 2112         | 2[6 <sup>3</sup> ] + [6 <sup>6</sup> ]                                 |
| 3,6 |     | Rutile (TiO <sub>2</sub> )     | T or Y, octahedron    | 2232         | 2[4 <sup>6</sup> .6 <sup>2</sup> ] + [6 <sup>2</sup> .8 <sup>2</sup> ] |
| 3,4 |     | Pt <sub>3</sub> O <sub>4</sub> | Triangle, square      | 2122         | 3[8 <sup>4</sup> ] + 2[8 <sup>6</sup> ]                                |
| 3,4 |     | Boracite                       | Triangle, tetrahedron | 2122         | [6 <sup>7</sup> ] + [6 <sup>3</sup> .8 <sup>6</sup> ]                  |
| 4,4 |     | PtS                            | Square, tetrahedron   | 2122         | [4 <sup>2</sup> .8 <sup>2</sup> ] + [8 <sup>4</sup> ]                  |
| 6,6 |     | NiAs                           | Prism, octahedron     | 2122         | [4 <sup>3</sup> ] + [4 <sup>9</sup> ]                                  |

CN is the coordination number of the vertices. LC refers to the symbol for invariant lattice complexes<sup>59</sup>. The transitivity is defined in the text, and the entries under ‘tiles’ are the face symbols of the tiles. The first six rows are the five regular and one quasi-regular tilings. For illustrations of these nets see refs 54 and 60.

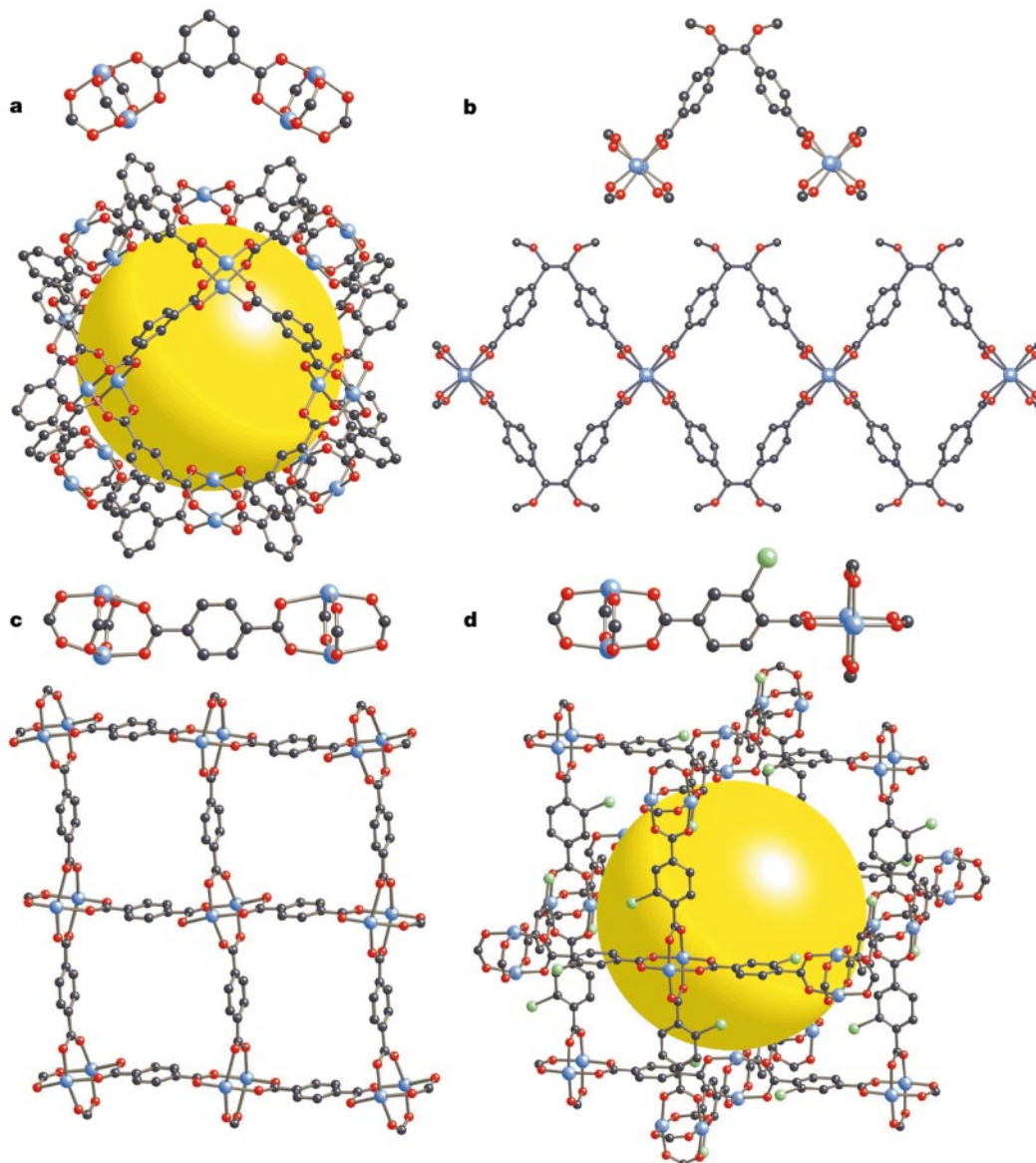
### A new class of porous materials

We have just argued that the rigidity and directionality of SBUs provide a means for the logical synthesis of a wide variety of frameworks. An important aspect of this approach is that the large sizes of SBUs inevitably lead to there being spaces within these structures where solvent and/or counter-ion guests reside. In many cases, such solids are interchangeably referred to as “porous” and “open-framework”, without definitive proof of porosity<sup>72</sup>. We prefer to categorize them according to the mobility of the species occupying their cavities. It is inaccurate to use the term ‘open framework’ to refer to any low-density or large-scale structure, because ‘open’ implies allowing passage. We note that in many cases this property is absent (or at least undetermined in the structure in question) because passage is hindered by solvent molecules that form strong interactions with the framework; as a result, such solvent molecules are necessary to maintain the crystal architecture. Thus we consider that ‘porous’ implies having pores, and pores

are, in the Oxford English Dictionary definition, a “minute... opening... through which fluids or gases can pass”. In addition, demonstration of porosity through gas sorption isotherms is necessary to prove permanent porosity. Indeed, we find that MOFs having structures based on inorganic SBUs are exceptionally porous with pore diameters and volumes which exceed those of the most porous and useful zeolites as shown in Fig. 4<sup>73–84</sup>.

Two important attributes of porous MOFs account for their structural stability and high porosity. First, they are essentially scaffoldings with all or most atoms on internal surfaces, which lead to the highest known surface areas ( $2,900\text{ m}^2\text{ g}^{-1}$ ) and pore volumes ( $1\text{--}2\text{ cm}^3\text{ g}^{-1}$ ). In other porous solids such as zeolites and molecular sieves the interior of the pores is largely composed of walls leading to relatively lower specific capacity for sorption.

Second, in contrast to frameworks of the metal bipyridine type, which are known to collapse upon removal of guests, MOFs constructed with oxide SBUs are far more robust, owing to the



**Figure 3** The control of dimensionality of linked paddle-wheel units by use of precise linker geometry. Shown are fragments of the assembled structures of MOP-161 (a), MOF-222 (Chae, H.K., Rowsell, J.R., O’Keeffe, M., Yaghi, O.M., to be published) (b),

MOF-242 (c) and MOF-101<sup>63</sup> (d). C, black; O, red; Br, green; metal, blue. Above each structure, one linker is shown joined to two paddle-wheel units. The yellow spheres are at the centres of large cavities in the structures of MOP-1 and MOF-101.



presence of strong metal–oxygen bonds. For example, in MOF-5 each link to an octahedral SBU involves two Zn–O bonds with energy of  $360 \text{ kJ mol}^{-1}$  per pair (based on the heat of formation of ZnO)<sup>85</sup> comparable to the C–C bond energy of  $358 \text{ kJ mol}^{-1}$  in diamond. The corresponding energy for two Cu–O bonds (as in linked paddle wheels) is  $372 \text{ kJ mol}^{-1}$ . In contrast, we estimate from the heats of formation of ammoniates such as  $\text{CuCl}\cdot 3\text{NH}_3$  and  $\text{CuCl}_2\cdot 2\text{NH}_3$  that the energy of the Cu(I)–N coordination bond is about  $55 \text{ kJ mol}^{-1}$  and the Cu(II)–N bond is about  $90 \text{ kJ mol}^{-1}$ : very much weaker, and indeed the weak link in metal-bipyridine type frameworks.

### Frameworks with adjustable metrics and properties

An important step forward has been the use of expanded links such as ECN which can be linked into MOF structures in which the pores can be functionalized by modifying  $\text{R}^{86}$ , and expanded by lengthening L (see Box 1). When assembled with Ag(I) this link gives networks based on the Si network in  $\text{ThSi}_2$  (Box 2). It is interesting to note that the unfunctionalized ( $\text{R} = \text{H}$ ) link also gives the same network. This finding indicates that the presence of functional groups ( $\text{R}$ ) does not alter the course of reticulation. In other words, all frameworks produced by this method are isorecticular (having the same network topology). These compounds have not been shown to have permanent porosity or rigid frameworks, which is not surprising as they are not based on vertices of clusters (SBUs), but rather on single Ag ion vertices, which do not provide for the prerequisite rigidity.

On the other hand, frameworks isorecticular with MOF-5 have been synthesized with functionalized structures and shown to have permanent porosity. Employing each of the links: *o*-Br-BDC,  $\text{R}_3$ -BDC,  $\text{R}_4$ -BDC,  $\text{R}_5$ -BDC,  $\text{R}_6$ -BDC,  $\text{R}_7$ -BDC, 2,6-NDC, BPDC, HPDC, PDC and TPDC instead of BDC yielded IRMOF-2 to IRMOF-16, including the non-interpenetrating structures of BPDC, HPDC, PDC and TPDC. Each member of the IRMOF series has been isolated and characterized by single-crystal X-ray

diffraction studies<sup>82</sup>. The key element in synthesizing this series was to use the acid form of the link under reaction conditions known to form the octahedral SBU. In this series, the per cent free volume in crystals varies in small increments (1 to 5%) starting from 55.8% ( $\text{R}_5$ -BDC) to an exceptional 91.1% (TPDC).

It is worth noting that the lowest per cent free volume obtained in this series exceeds that found in some of the most open zeolites (Fig. 4) in which the free space is 45–50% of the crystal volume. In fact, the fraction of free space in crystals of the most expanded member (TPDC) of this series has only been achievable in non-crystalline porous systems such as  $\text{SiO}_2$  xerogels and aerogels<sup>87</sup>. Furthermore, the calculated crystal densities (in the absence of guests) of these materials also vary in small increments (around 0.1) in the range  $1.00 \text{ g cm}^{-3}$  ( $\text{R}_5$ -BDC) to  $0.21 \text{ g cm}^{-3}$  (TPDC) having densities, which are the lowest reported for any crystalline material. In comparison, the density of Li metal is  $0.56 \text{ g cm}^{-3}$ . The ability to control and incrementally vary the metrics in this series has important technological implications. It has already been shown that the specific uptake of fuel gases such as methane in one member ( $\text{R}_6$ -BDC) of the series is the highest known at present and comes close to meeting Department of Energy guidelines for use of natural gas in mobile applications.

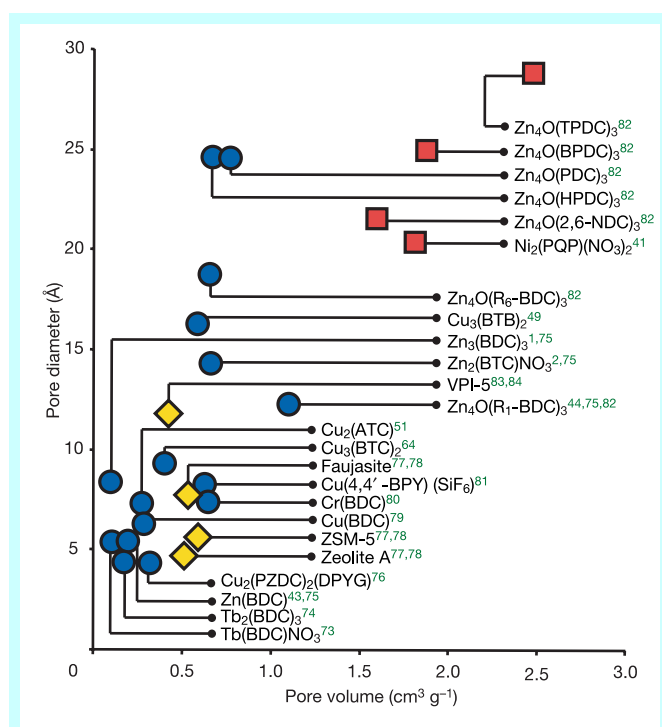
In cognate work an isorecticular series based on lamellar G and S (guanidine-sulphonate) hydrogen-bonded materials has been achieved by varying the sulphonate functionality from DS-PGBDS (Box 1). Although these frameworks do not have permanent porosity as in MOFs, they provide advantages for selective crystallization-based separations<sup>88</sup>.

In addition to providing rigidity and structure information, SBUs can also impart physical properties to the framework. In the examples based on the paddle wheel and other binuclear clusters, the axial ligands are weakly bound to the porous framework, allowing their removal by heating or evacuation. The resulting open metal sites act as molecular imprints and site-isolated units<sup>51</sup>, exhibiting highly selective binding of organic molecules and, in the case of Tb(III), luminescent behaviour that has formed a basis for sensing small molecules<sup>74</sup>.

Although porosity is almost an inevitable outcome of the SBU approach, other MOFs that have not been examined for their porosity have shown other interesting properties. Acentric SBUs can be assembled into MOFs having nonlinear optical properties, as illustrated for PCA-1<sup>89</sup>. Here, a triangular SBU is linked by PCA-1 to give a decorated honeycomb-type framework. Similar frameworks have been designed using paramagnetic metal ions to produce ferromagnetic MOFs: links such as OX and OXM serve as magnetic coupling bridges in 2D ferromagnetic frameworks<sup>90,91</sup>. Recently, 3D magnetic frameworks have also been targeted using Mn(II), Fe(II), Co(II), Ni(II), Cu(II) and Cr(III) metal ions and  $\text{CN}^{92}$ ,  $\text{C}(\text{CN})_3^{93}$  and  $\text{Re}_6\text{S}_8(\text{CN})_6^{94}$  links to produce rutile and Prussian-blue type solids. A mixed-valence conductive framework based on BDC has been constructed from V(III/IV)<sup>80</sup>, in which infinite vanadium oxide SBUs lead to the same  $\text{SrAl}_2$  topology found for MOF-69<sup>95</sup>.

### Logic of synthesis by design

On the conceptual level, the underlying process for logical synthesis of MOFs starts with the designer choosing a specific target network, which is then deconstructed into its component geometric units. Molecular building blocks with the same geometry as those of the units are then assembled into a MOF structure that has the target network topology. The value of this approach is not limited to its utility in accessing a vast number of materials with predetermined structures: it also allows the designer to develop a deconstruction strategy for a target network. Such a deconstruction strategy becomes indispensable when designing non-default networks. This aspect is also addressed by reticular synthesis as shown in Fig. 5 for two basic networks; diamond and lonsdaleite ('hexagonal diamond')<sup>54,60</sup>.

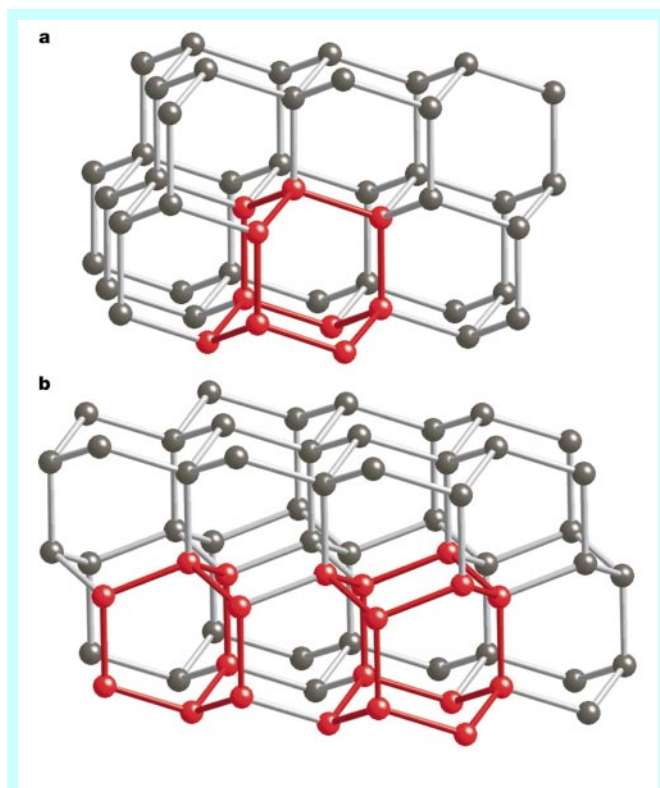


**Figure 4** The porosity of MOFs compared to zeolites. Blue circles, sorption and crystal MOF structures; red squares, crystal MOF structures; yellow diamonds, some typical zeolites. Reference numbers (superscripted) are shown in green.

The designer of a material having these topologies would recognize that the simplest deconstruction scheme yields tetrahedral building blocks in both cases. Thus the assembly of molecules with tetrahedral shape would be expected to yield at least one of these two possible arrangements. However, in practice most structures (discussed above) obtained from simple tetrahedral building blocks are based on the cubic diamond network (Fig. 5a), which corresponds to the simplest, highest-symmetry structure (it is the only regular tetrahedral net) and unless the molecular building blocks contain information to the contrary, it is the default reticulation. Reticulations based on the lonsdaleite topology are rare.

We therefore investigated how reticular synthesis could be used to access non-default structures such as lonsdaleite. It is necessary to deconstruct its structure into more elaborate units that express structural features unique to its structure. Although the two structures are composed of fused six-membered rings, all such rings in diamond have the 'chair' conformation, whereas those in the hexagonal form assume both chair and 'boat' conformations as illustrated in Fig. 5. Thus it seems reasonable that to make frameworks based on lonsdaleite, the network should be deconstructed into units of fused chair and boat rings rather than of individual tetrahedra. In this way, the assembly of the corresponding molecular building blocks would yield exclusively the target network. We note that the above discussion does not preclude the possibility of achieving lonsdaleite from smaller units, if the structure formation were driven by the templating effect of included guest molecules<sup>96</sup> as described in the next section.

This treatment can be universally applied to the design of crystalline molecular arrays of any dimensionality. We emphasize that the achievement of complex low-symmetry topologies by design requires the construction of complex molecular building blocks coded specifically for those topologies.



**Figure 5** The deconstruction of diamond and lonsdaleite. **a**, diamond; **b**, lonsdaleite ('hexagonal diamond'). The units outlined in red are the natural tiles for these structures.

## Complexity and non-default structures

The discussion so far has been limited to the design of frameworks in which the geometry of the links and/or vertices largely determines the outcome. In the formation of neutral frameworks, such as those of metal carboxylates, the role of the solvent in default structures is largely a space-filling one, although their function as structure-directing agents (templates) becomes more important in the design of networks where the building blocks have more than one possible default structure<sup>96</sup>.

In particular, assembly of charged frameworks with counter ions requires careful matching of the charge, size and volume of counter ion to the charge per unit area of the target framework. This was illustrated in the assembly of T3 and T4 indium sulphide super-tetrahedra, whose tetrahedral symmetry leads not only to the default diamond-like frameworks<sup>97</sup>, but also to the slightly less symmetric structures such as that of sodalite and the boron network in  $\text{CrB}_4$ <sup>98</sup>. Here it is important to consider both the size and volume of the organic ammonium cations and the charge they provide per unit area of the prospective charged framework. We expect that similar factors must be considered to explain the formation of sodalite, feldspar and quartz MOF networks from tetrahedral  $\text{Cu(I)}$  and HPMO<sup>99</sup>, PRM<sup>100</sup> and 2,2'-bpy<sup>39</sup>, respectively. This analysis indicates that for some target non-default networks, elaborate units similar to those suggested for the reticular synthesis of lonsdaleite may not be the only option for their design.

## Future prospects

Even with only the structures that are based on default networks, the immense wealth and scope of reticular chemistry has become apparent. In particular, design of the structural skeleton coupled with the ability to control its chemical functionalization and adjustment of its metric dimensions present exciting prospects. We have described here a conceptual approach to this subject, which we believe has proved useful, and which will continue to guide the designer.

Specific future challenges include the ability to design sophisticated molecular building blocks that would act as carriers of structural and functional information to be expressed in a specific target material. This will be of great benefit to many emerging technologies involving molecular electronics, miniature fuel cells, chiral catalysis and materials with hybrid properties. □

doi:10.1038/nature01650.

- Stein, A., Keller, S. W. & Mallouk, T. E. Turning down the heat: Design and mechanism in solid-state synthesis. *Science* **259**, 1558–1563 (1993).
- Yaghi, O. M., O'Keeffe, M. & Kanatzidis, M. G. Design of solids from molecular building blocks: golden opportunities for solid state chemistry. *J. Solid State Chem.* **152**, 1–2 (2000).
- Yaghi, O. M., Li, H., Davis, C., Richardson, D. & Groy, T. L. Synthetic strategies, structure patterns, and emerging properties in the chemistry of modular porous solids. *Acc. Chem. Res.* **31**, 474–484 (1998).
- Batten, S. T. & Robson, R. Interpenetrating nets: Ordered, periodic entanglement. *Angew. Chem. Int. Edn Engl.* **37**, 1460–1494 (1998).
- Férey, G. Building units design and scale chemistry. *J. Solid State Chem.* **152**, 37–48 (2000).
- Kitagawa, S. & Kondo, M. Functional micropore chemistry of crystalline metal complex-assembled compounds. *Bull. Chem. Soc. Jpn* **71**, 1739–1753 (1998).
- Yaghi, O. M., O'Keeffe, M. & Kanatzidis, M. G. Special issue on the design of solids from molecular building blocks: golden opportunities for solid state chemistry. *J. Solid State Chem.* **152**, 1–321 (2000).
- Eddaoudi, M. *et al.* Modular chemistry: Secondary building units as a basis for the design of highly porous and robust metal-organic carboxylate frameworks. *Acc. Chem. Res.* **34**, 319–330 (2001).
- Yaghi, O. M., Sun, Z., Richardson, D. A. & Groy, T. L. Directed transformation of molecules to solids: Synthesis of microporous sulfide from molecular germanium sulfide cages. *J. Am. Chem. Soc.* **116**, 807–808 (1994).
- Corey, E. J. Retrosynthetic thinking-essentials and examples. *Chem. Soc. Rev.* **17**, 111–133 (1988).
- Lehn, J. M. Supramolecular chemistry-scope and perspectives. *Chem. Scr.* **28**, 237–262 (1988).
- Discussion 1: Innovation in Crystal Engineering *CrystEngComm* [online]; available at (<http://www.rsc.org/is/journals/current/crystengcomm/ceced.htm>) (August 2002).
- Moulton, B. & Zaworotko, M. J. From molecules to crystal engineering: Supramolecular isomerism and polymorphism in network solids. *Chem. Rev.* **101**, 1629–1658 (2001).
- Braga, D., Grepioni, F. & Desiraju, G. R. Crystal engineering and organometallic architecture. *Chem. Rev.* **98**, 1375–1406 (1998).
- Wells, A. F. *Structural Inorganic Chemistry*, (Oxford University Press, New York, 1984).

16. Dunbar, K. R. & Heintz, R. A. Chemistry of transition metal cyanide compounds: Modern perspectives. *Prog. Inorg. Chem.* **45**, 283–391 (1997).
17. Gramaccioli, C. M. Crystal structure of zinc glutamate dihydrate. *Acta Crystallogr.* **21**, 600–605 (1966).
18. Okada, K., Kay, M. I., Cromer, D. T. & Almodovar, I. Crystal structure by neutron diffraction and antiferroelectric phase transition in copper formate tetrahydrate. *J. Chem. Phys.* **44**, 1648–1653 (1966).
19. Jarvis, J. A. J. Crystal structure of a complex of cupric chloride and 1, 2, 4-triazole. *Acta Crystallogr.* **15**, 964–966 (1962).
20. Sterling, C. Crystal structure of weddellite. *Science* **146**, 518–519 (1964).
21. Robl, C. Water clustering in the zeolite-like channel structure of  $\text{Na}_2\text{Zn}(\text{C}_6\text{H}_2(\text{COO})_4)_9\text{H}_2\text{O}$ . *Mater. Res. Bull.* **27**, 99–107 (1992).
22. Weiss, A., Riegler, E., Alt, I., Bohme, H. & Robl, C. Transition metal squarates 1. Chain structures  $\text{M}(\text{C}_4\text{O}_4)\cdot 4\text{H}_2\text{O}$ . *Z. Naturforsch. B.* **41**, 18–24 (1986).
23. Kinoshita, Y., Matsubara, I., Higuchi, T. & Saito, Y. The crystal structure of bis (adiponitrile)copper(I) nitrate. *Bull. Chem. Soc. Jpn* **32**, 1221–1226 (1959).
24. Aumüller, A. et al. A radical anion salt of 2, 5-dimethyl-N, N'-dicyanoquinonediimine with extremely high electrical conductivity. *Angew. Chem. Int. Edn Engl.* **25**, 740–741 (1986).
25. Hoskins, B. F. & Robson, R. Infinite polymeric frameworks consisting of three dimensionally linked rod-like segments. *J. Am. Chem. Soc.* **111**, 5962–5964 (1989).
26. Kim, J. et al. Assembly of metal-organic frameworks from large organic and inorganic secondary building units: new examples and simplifying principles for complex structures. *J. Am. Chem. Soc.* **123**, 8239–8247 (2001).
27. Brimah, A. K. et al. Towards organometallic zeolites: spontaneous self-assembly of  $\text{Et}_3\text{SnCN}$ ,  $\text{CuCN}$  and  $(^t\text{Bu}_4\text{N})\text{CN}$  to supramolecular  $[(^t\text{Bu}_4\text{N})(\text{Et}_3\text{Sn})_2\text{Cu}(\text{CN})_4]$ . *J. Organometall. Chem.* **475**, 85–94 (1994).
28. Evans, O. R., Wang, Z. Y., Xiong, R. G., Foxman, B. M. & Lin, W. B. Nanoporous interpenetrated metal-organic diamondoid networks. *Inorg. Chem.* **38**, 2969–2973 (1999).
29. Yaghi, O. M., Richardson, D. A., Li, G., Davis, C. E. & Groy, T. L. Open-framework solids with diamond-like structures prepared from clusters and metal-organic building blocks. *Mater. Res. Soc. Symp. Proc.* **371**, 15–19 (1995).
30. MacGillivray, L. R., Subramanian, S. & Zaworotko, M. J. Interwoven two- and three-dimensional coordination polymers through self-assembly of Cu (I) cations with linear bidentate ligands. *Chem. Commun.*, 1325–1326 (1994).
31. Hirsch, K. A., Venkataraman, D., Wilson, S. R., Moore, J. S. & Lee, S. Crystallization of 4, 4'-biphenyldicarbonitrile with silver (I) salts - a change in topology concomitant with change in counterion leading to a nine diamond network. *Chem. Commun.*, 2199–2200 (1995).
32. Carlucci, L. et al. Self-assembly of a three-dimensional network from two dimensional layers via metallic spacers: the (3,4)-connected frame of  $[\text{Ag}_3(\text{hmt})_2][\text{ClO}_4]_2\cdot 2\text{H}_2\text{O}$  (hmt = hexamethylenetetramine). *Chem. Commun.*, 631–632 (1997).
33. Klein, C., Graf, E., Hosseini, M. W. & De Cian, A. Design and structural analysis of interpenetrated 3-D coordination networks formed from self-assembly using tetrapyrrolinecyclophane and silver cations. *New J. Chem.* **25**, 207–209 (2001).
34. Simard, M., Su, D. & Wuest, J. D. Use of hydrogen bonds to control molecular aggregation. Self-assembly of three-dimensional networks with large chambers. *J. Am. Chem. Soc.* **113**, 4696–4698 (1991).
35. Ermer, O. Fivefold-diamond structure of adamantane-1,3,5,7-tetracarboxylic acid. *J. Am. Chem. Soc.* **110**, 3747–3754 (1988).
36. Blake, A. J., Champness, N. R., Hubbertsey, P., Li, W., Withersby, M. A. & Schroder, M. Inorganic crystal engineering using self-assembly of tailored building-blocks. *Coord. Chem. Rev.* **183**, 117–138 (1999).
37. Carlucci, L., Ciani, G., Macchi, P. & Proserpio, D. M. An unprecedented triply interpenetrated chiral network of 'square planar' metal centres from the self assembly of copper(II) nitrate and 1,2-bis(4-pyridyl)ethyne. *Chem. Commun.*, 1837–1838 (1998).
38. Carlucci, L., Ciani, G., Proserpio, D. M. & Rizzato, S. New examples of self-catenation in three dimensional co-ordination networks. *J. Chem. Soc. Dalton Trans.*, 3821–3827 (2000).
39. Carlucci, L., Ciani, G., Proserpio, D. M. & Rizzato, S. Interlinked molecular squares with  $[\text{Cu}(2,2' - \text{bipy})]^{2+}$  corners generating a three-dimensional network of unprecedented topological type. *Chem. Commun.*, 1198–1199 (2001).
40. Evans, O. R. & Lin, W. Crystal engineering in NLO materials based on metal-organic coordination networks. *Acc. Chem. Res.* **35**, 511–522 (2002).
41. Biradha, K., Hongo, Y. & Fujita, M. Open square-grid coordination polymers of the dimension  $20 \times 20 \text{ \AA}$ : remarkably stable and crystalline solids even after guest removal. *Angew. Chem. Int. Edn Engl.* **39**, 3843–3845 (2000).
42. Yaghi, O. M., Li, H. & Groy, T. L. A molecular railroad with large pores: synthesis and structure of  $\text{Ni}(4,4' - \text{bpy})_{2.5}(\text{H}_2\text{O})_2(\text{ClO}_4)_2 \cdot 1.5(4,4' - \text{bpy}) \cdot 2\text{H}_2\text{O}$ . *Inorg. Chem.* **36**, 4292–4293 (1997).
43. Li, H., Eddaoudi, M., Groy, T. L. & Yaghi, O. M. Establishing microporosity in open metal-organic frameworks: gas sorption isotherms for Zn(BDC) (BDC = 1,4-benzenedicarboxylate). *J. Am. Chem. Soc.* **120**, 8571–8572 (1998).
44. Li, H., Eddaoudi, M., O'Keeffe, M. & Yaghi, O. M. Design and synthesis of an exceptionally stable and highly porous metal-organic framework. *Nature* **402**, 276–279 (1999).
45. Yaghi, O. M., Davis, C. E., Li, G. & Li, H. Selective guest binding by tailored channels in a 3-D porous zinc(II) benzenetricarboxylate network. *J. Am. Chem. Soc.* **119**, 2861–2868 (1997).
46. Barthelot, K., Riou, D. & Férey, G.  $[\text{V}^{\text{III}}(\text{H}_2\text{O})]_3\text{O}(\text{O}_2\text{CC}_6\text{H}_4\text{CO}_2)_3 \cdot (\text{Cl}, 9\text{H}_2\text{O})$  (MIL-59): A rare example of vanadocarboxylate with a magnetically frustrated three-dimensional hybrid framework. *Chem. Commun.*, 1492–1493 (2002).
47. Yang, S. Y., Long, L. S., Jiang, Y. B., Huang, R. B. & Zheng, L. S. An exceptionally stable metal-organic framework constructed from the  $\text{Zn}_6(\text{SiO}_4)_4$  core. *Chem. Mater.* **14**, 3229–3231 (2002).
48. Eddaoudi, M. et al. Geometric requirements and examples of important structures in the assembly of square building blocks. *Proc. Natl. Acad. Sci. USA* **99**, 4900–4904 (2002).
49. Chen, B., Eddaoudi, M., Hyde, S. T., O'Keeffe, M. & Yaghi, O. M. Interwoven metal-organic framework on a periodic minimal surface with extra-large pores. *Science* **291**, 1021–1023 (2001).
50. Diskin-Posner, Y., Dahal, S. & Goldberg, I. Crystal engineering of metalloporphyrin zeolite analogues. *Angew. Chem. Int. Edn Engl.* **39**, 1288–1292 (2000).
51. Chen, B. et al.  $\text{Cu}_2(\text{ATC})\cdot 6\text{H}_2\text{O}$ : Design of open metal sites in porous metal-organic crystals (ATC: 1,3,5,7-adamantane tetracarboxylate). *J. Am. Chem. Soc.* **122**, 11559–11560 (2000).
52. Chae, H. K. et al. Tertiary building units: Synthesis, structure, and porosity of a metal-organic dendrimer framework (MODF-1). *J. Am. Chem. Soc.* **123**, 11482–11483 (2001).
53. O'Keeffe, M. & Brese, N. E. Uninodal 4-connected nets I: Nets without 3- or 4-rings. *Acta Crystallogr. A* **48**, 663–669 (1992).
54. O'Keeffe, M., Eddaoudi, M., Li, H., Reineke, T. & Yaghi, O. M. Frameworks for extended solids: geometrical design principles. *J. Solid State Chem.* **152**, 3–20 (2000).
55. Delgado Friedrichs, O., Dress, A. W. M., Huson, D. H., Klinowski, J. & Mackay, A. L. Systematic enumeration of crystalline networks. *Nature* **400**, 644–647 (1999).
56. O'Keeffe, M. Tiling by numbers. *Nature* **400**, 617–618 (1999).
57. Delgado Friedrichs, O. & Huson, D. H. 4-regular vertex-transitive tilings of  $\text{E}^3$ . *Discr. Comput. Geom.* **24**, 279–292 (2000).
58. Fischer, W. & Koch, E. in *International Tables For Crystallography A* (ed. Hahn, Th.) Ch. 14 (Kluwer, Dordrecht, 1983).
59. Delgado Friedrichs, O., O'Keeffe, M. & Yaghi, O. M. Three-periodic nets and tilings: regular and quasiregular nets. *Acta Crystallogr. A* **59**, 22–27 (2003).
60. O'Keeffe, M. & Hyde, B. G. *Crystal structures, I. Patterns and symmetry* (Mineralogical Society of America, Washington, DC, 1996).
61. Eddaoudi, M. et al. Porous metal-organic polyhedra: 25 angstrom cuboctahedron constructed from 12  $\text{Cu}_2(\text{CO}_2)_4$  paddle wheel building blocks. *J. Am. Chem. Soc.* **123**, 4368–4369 (2001).
62. Moulton, B., Lu, J., Mondal, A. & Zaworotko, M. Z. Nanoballs: nanoscale faceted polyhedra with large windows and cavities. *Chem. Commun.*, 863–864 (2001).
63. Eddaoudi, M., Kim, J., O'Keeffe, M. & Yaghi, O. M.  $\text{Cu}_2[\text{o}-\text{Br}-\text{C}_6\text{H}_3(\text{CO}_2)_2]_2(\text{H}_2\text{O})_2(\text{DMF})_2(\text{H}_2\text{O})_2$ : A framework deliberately designed to have the NbO structure type. *J. Am. Chem. Soc.* **124**, 376–377 (2002).
64. Chu, S. S. Y., Lo, S. M. F., Charmant, J. P. H., Orpen, A. G. & Williams, I. D. A chemically functionalizable nanoporous material  $\text{Cu}_3(\text{TMA})_2(\text{H}_2\text{O})_3$ . *Science* **283**, 1148–1150 (1999).
65. Kepert, C. J., Prior, T. J. & Rosseinsky, M. J. A versatile family of interconvertible microporous chiral molecular frameworks: the first example of ligand control of network chirality. *J. Am. Chem. Soc.* **122**, 5158–5168 (2000).
66. Seo, J. S. et al. A homochiral metal-organic porous material for enantioselective separation and catalysis. *Nature* **404**, 982–986 (2000).
67. Fujita, M. et al. On the structure of transition-metal-linked molecular squares. *Chem. Commun.*, 1535–1536 (1996).
68. Stang, P. T. & Olenyuk, B. Self-assembly, symmetry, and molecular architecture: coordination as the motif in the rational design of supramolecular metallocyclic polygons and polyhedra. *Acc. Chem. Res.* **30**, 502–518 (1997).
69. MacGillivray, L. R. & Atwood, J. L. A chiral spherical molecular assembly held together by 60 hydrogen bonds. *Nature* **389**, 469–472 (1997).
70. Caulder, D. L. & Raymond, K. N. Supermolecules by design. *Acc. Chem. Res.* **32**, 975–982 (1999).
71. Cotton, F. A., Lin, C. & Murillo, C. A. Supramolecular arrays based on dimetal building units. *Acc. Chem. Res.* **34**, 759–771 (2001).
72. Gregg, S. J. & Sing, K. S. W. *Adsorption, Surface Area and Porosity* (Academic Press, New York, 1982).
73. Reineke, T. M., Eddaoudi, M., O'Keeffe, M. & Yaghi, O. M. A microporous lanthanide-organic framework. *Angew. Chem. Int. Edn Engl.* **38**, 2590–2594 (1999).
74. Reineke, T., Eddaoudi, M., Fehr, M., Kelley, D. & Yaghi, O. M. From condensed lanthanide coordination solids to microporous frameworks having accessible metal sites. *J. Am. Chem. Soc.* **121**, 1651–1657 (1999).
75. Eddaoudi, M., Li, H. & Yaghi, O. M. Highly porous and stable metal-organic frameworks: Structure design and sorption properties. *J. Am. Chem. Soc.* **122**, 1391–1397 (2000).
76. Kitaura, R., Fujimoto, K., Noro, S., Kondo, M. & Kitagawa, S. A pillared-layer coordination polymer displaying hysteretic sorption:  $[\text{Cu}_2(\text{pzdc})_2(\text{dpyg})]_n$  (pzdc = pyrazine-2,3-dicarboxylate; dpyg = 1,2-Di(4-pyridyl)-glycol). *Angew. Chem. Int. Edn Engl.* **41**, 133–135 (2002).
77. Szołtak, R. *Molecular Sieves: Principles of Synthesis and Identification* (Van Nostrand Reinhold, New York, 1989).
78. Bhatia, S. *Zeolite Catalysis: Principles and Applications* (CRC Press, Boca Raton, Florida, 1990).
79. Seki, K., Takamizawa, S. & Mori, W. Design and gas adsorption property of a three-dimensional coordination polymer with a stable and highly porous framework. *Chem. Lett.*, 332–333 (2001).
80. Millange, F., Serre, C. & Férey, G. Synthesis, structure determination and properties of MIL-53as and MIL-53ht: the first  $\text{Cr}^{\text{III}}$  hybrid inorganic-organic microporous solids:  $\text{Cr}^{\text{III}}(\text{OH})\cdot [\text{O}_2\text{C}-\text{C}_6\text{H}_4-\text{CO}_2]_2\cdot [\text{HO}_2-\text{C}_6\text{H}_4-\text{CO}_2]_x$ . *Chem. Commun.*, 822–823 (2002).
81. Noro, S., Kitagawa, S., Kondo, M. & Seki, M. A new methane adsorbent, porous coordination polymer  $[\{\text{CuSiF}_6(4,4' - \text{bipyridine})_2\}]_n$ . *Angew. Chem. Int. Edn Engl.* **39**, 2082–2084 (2000).
82. Eddaoudi, M. et al. Systematic design of pore size and functionality in isorecticular MOFs and their application in methane storage. *Science* **295**, 469–472 (2002).
83. Davis, M. E., Montes, C., Hathaway, P. E., Arhancet, J. P., Hasha, D. L. & Garces, J. M. Physicochemical properties of VPI-5. *J. Am. Chem. Soc.* **111**, 3919–3924 (1989).
84. Davis, M. E. Ordered porous materials for emerging applications. *Nature* **417**, 813–821 (2002).
85. O'Keeffe, M. & Stuart, J. A. Bond energies in solid oxides. *Inorg. Chem.* **22**, 177–179 (1983).
86. Kiang, Y. H., Gardner, G. B., Lee, S., Xu, Z. & Lobkovsky, E. B. Variable pore size, variable chemical functionality, and an example of reactivity within porous phenylacetylene silver salts. *J. Am. Chem. Soc.* **121**, 8204–8215 (1999).
87. Husing, N. & Schubert, U. Aerogels airy materials: Chemistry, structure, and properties. *Angew. Chem. Int. Edn Engl.* **37**, 23–45 (1998).
88. Holman, K. T., Pivovar, A. M., Swift, J. A. & Ward, M. D. Metric engineering of soft molecular host frameworks. *Acc. Chem. Res.* **34**, 107–118 (2001).
89. Lin, W. B., Wang, Z. & Ma, L. A novel octupolar metal-organic NLO material based on chiral 2D-coordination network. *J. Am. Chem. Soc.* **121**, 11249–11250 (1999).
90. Farrell, R. P., Hambley, T. W. & Lay, P. A. A new class of layered microporous materials: crystal structure of disodium pentakis(trimethylphenylammonium) bis[tris(oxalato(2))-chromate(III)] chloride pentahydrate. *Inorg. Chem.* **34**, 757–758 (1995).
91. Kahn, O. Chemistry and physics of supramolecular magnetic materials. *Acc. Chem. Res.* **33**, 647–657 (2000).



92. Sato, O., Iyoda, T., Fujishima, A. & Hashimoto, K. Electrochemically tunable magnetic phase transition in a high- $T_c$  chromium cyanide thin film. *Science* **271**, 49–51 (1996).
93. Manson, J. L., Campana, C. & Miller, J. S. Interpenetrating three-dimensional rutile-like frameworks. Crystal structure and magnetic properties  $\text{Mn}^{\text{II}}[\text{C}(\text{CN})_3]_2$ . *Chem. Commun.*, 251–252 (1998).
94. Bennett, M. V., Beauvais, L. G., Shores, M. P. & Long, J. R. Expanded prussian blue analogues incorporating  $[\text{Re}_6\text{Se}_8(\text{CN})_6]^{3-/4-}$  clusters: adjusting porosity via charge balance. *J. Am. Chem. Soc.* **123**, 8022–8032 (2001).
95. Rosi, N. L. *et al.* Infinite secondary building units and forbidden catenation in metal-organic frameworks, N. L. *Angew. Chem. Int. Edn Engl.* **41**, 284–287 (2001).
96. Kitazawa, T., Kikuyama, T., Takeda, M. & Iwamoto, T. Silica-mimetic polymorphism of the  $\text{Cd}(\text{CN})_2$  host lattice depending on the guest G in  $\text{Cd}(\text{CN})_{2-x}\text{G}$  clathrates. *J. Chem. Soc. Dalton Trans.* **22**, 3715–3720 (1995).
97. Li, H., Eddaoudi, M., Laine, A., O’Keeffe, M. & Yaghi, O. M. Noninterpenetrating indium sulfide supertetrahedral cristobalite framework. *J. Am. Chem. Soc.* **121**, 6096–6097 (2001).
98. Li, H., Laine, A., O’Keeffe, M. & Yaghi, O. M. Supertetrahedral sulfide crystals with giant cavities and channels. *Science* **283**, 1145–1147 (1999).
99. Tabares, L. C., Navarro, J. A. R. & Salas, J. M. Cooperative guest inclusion by a zeolite analogue coordination polymer: sorption behaviour with gases and amine and group 1 metal salts. *J. Am. Chem. Soc.* **123**, 383–387 (2001).
100. Keller, S. W. An acentric, three dimensional coordination polymer: synthesis and structure of  $[\text{Cu}(\text{pyrimidine})_2]\text{BF}_4$ . *Angew. Chem. Int. Edn Engl.* **36**, 247–248 (1997).

**Acknowledgements** O.M.Y. and M.O’K. thank the NSF and the DOE (O.M.Y.) for their support of research in their respective laboratories on the subject of this contribution.

**Correspondence** and requests for materials should be addressed to O.M.Y. (oyaghi@umich.edu).

# Adaptive evolution drives divergence of a hybrid inviability gene between two species of *Drosophila*

Daven C. Presgraves\*, Lakshmi Balagopalan†, Susan M. Abmayr† & H. Allen Orr\*

\* Department of Biology, University of Rochester, Rochester, New York 14627, USA

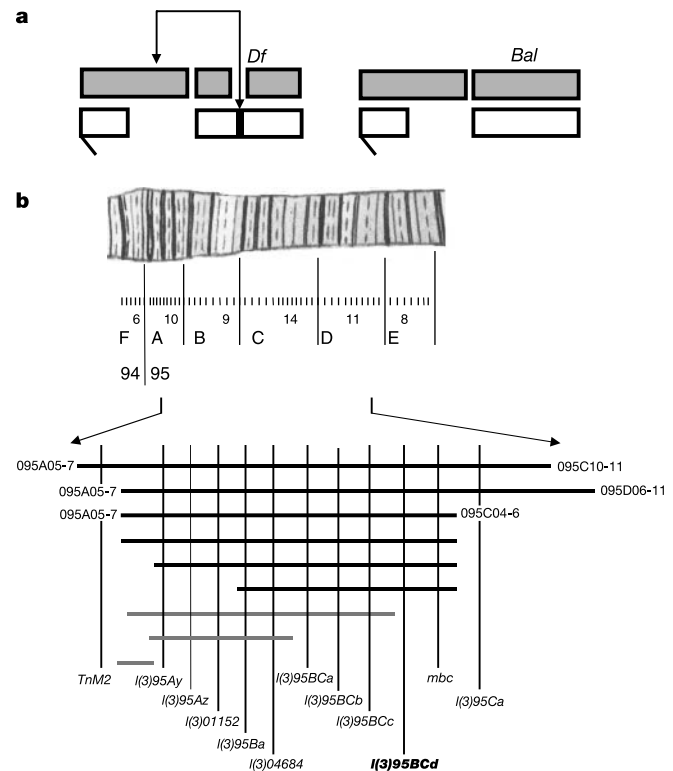
† Department of Biochemistry and Molecular Biology, The Pennsylvania State University, 459 North Frear Lab, University Park, Pennsylvania 16802, USA

**Speciation—the splitting of one species into two—occurs by the evolution of any of several forms of reproductive isolation between taxa, including the intrinsic sterility and inviability of hybrids. Abundant evidence shows that these hybrid fitness problems are caused by incompatible interactions between loci: new alleles that become established in one species are sometimes functionally incompatible with alleles at interacting loci from another species. However, almost nothing is known about the genes involved in such hybrid incompatibilities or the evolutionary forces that drive their divergence. Here we identify a gene that causes epistatic inviability in hybrids between two fruitfly species, *Drosophila melanogaster* and *D. simulans*. Our population genetic analysis reveals that this gene—which encodes a nuclear pore protein—evolved by positive natural selection in both species' lineages. These results show that a lethal hybrid incompatibility has evolved as a by-product of adaptive protein evolution.**

One of the long-standing goals of speciation research is to establish the molecular identities, functions and evolutionary histories of the genes that cause hybrid sterility and inviability<sup>1,2</sup>. So far, however, only three putative hybrid incompatibility genes have been identified<sup>3–5</sup>. One reason for this paucity of molecular data is that the characters of interest—hybrid sterility and inviability—are by their nature barriers to crossing. These characters are therefore often refractory to many classical genetic approaches<sup>6</sup>. Historically, this problem has proved especially serious in one of our best genetic model organisms, *D. melanogaster*, which is completely reproductively isolated from its closest relatives. All hybrids between *D. melanogaster* and its sibling species are sterile or inviable<sup>7,8</sup> (but see refs 9, 10). Thus, the impressive complement of genetic and molecular tools available in *D. melanogaster* has not been fully brought to bear on the genetics of speciation.

To address this problem, we previously carried out a systematic screen that makes use of the genetic tools of *D. melanogaster*<sup>11</sup>. This screen takes advantage of the fact that most hybrid incompatibilities involve epistatically interacting alleles<sup>12,13</sup> that act as recessives in species hybrids<sup>13,14</sup>. Crosses between *D. melanogaster* females and *D. simulans* males normally produce only hybrid daughters—hybrid males die at the larval–pupal transition<sup>7</sup>. Our crossing scheme involves first rescuing hybrid males from this hybrid incompatibility and then exposing them to other potential hybrid incompatibilities. We cross *D. melanogaster* females heterozygous for an autosomal deficiency (a small chromosomal deletion; *Df*) and a balancer (*Bal*) chromosome to *D. simulans* males carrying *Lethal hybrid rescue* (*Lhr*), a mutation that restores the viability of ordinarily lethal hybrid males<sup>15</sup>. We are interested in the relative viability of *Df* hybrid males that are simultaneously hemizygous for a small region of the *D. simulans* autosomal genome and the *D. melanogaster* X chromosome (Fig. 1a; see Methods). These males are forced to develop using only *D. simulans* (*sim*) alleles at the autosomal loci deleted and only *D. melanogaster* (*mel*) alleles at all X-linked loci (Fig. 1a), allowing us to detect any recessive alleles on the *sim* autosomes that are involved in hybrid lethal interactions with recessive alleles on the *mel* X chromosome. By screening ~70% of the *D. simulans* autosomal genome, we identified 20 small regions, each of which is capable of causing complete hybrid inviability<sup>11</sup>.

Here we present the fine-scale genetic, molecular and evolution-



**Figure 1** Deficiency mapping hybrid lethality. **a**, Schematic of hybrid male genotypes tested, with the sex chromosomes (left, Y chromosome with hook) and one autosome (right) shown. Grey chromosomes, *D. melanogaster*; white, *D. simulans*. Hybrid males are produced from the cross of *Df/Bal D. melanogaster* females to *Lhr D. simulans* males. Black, autosomal region causing hybrid lethality by its incompatibility with the *D. melanogaster* X chromosome. **b**, Interspecific complementation tests using deficiencies (horizontal lines) and loss-of-function mutations at loci (vertical lines) in cytological region 95 of chromosome arm 3R. Grey, complements hybrid lethality; black, fails to complement hybrid lethality. Deficiencies tested are, in order, from the top: *Df(3R)mbc-30*; *Df(3R)mbc-R1*; *Df(3R)mbc-BG1*; *Df(3R)mbc-15A*; *Df(3R)nau-9*; *Df(3R)mbc-F5.3*; *Df(3R)nau-4a*; *Df(3R)CA*; *Df(3R)nau-11a4* (see refs 45, 46). Only alleles of *I(3)95BCd* fail to complement hybrid lethality.

ary analysis of the first of the hybrid inviability genes that we have been able to identify.

### Nup96 causes inviability in species hybrids

We mapped a hybrid inviability gene in hybrid males using nine overlapping deficiencies from *D. melanogaster* to cytological region 95AB on chromosome 3R (Fig. 1b). These crosses narrowed hybrid inviability to just two possible complementation groups: *l(3)95BCd* and *myoblast city (mbc)*. We nevertheless tested presumed loss-of-function mutations at 12 loci that span the region, singly testing individual *mel* mutations for their ability to uncover hybrid lethality when heterozygous with the *sim* wild-type allele. We found that mutant alleles at only one locus failed to complement the *sim* hybrid lethal factor, *l(3)95BCd* (Fig. 1b). To confirm that the *sim* wild-type allele of *l(3)95BCd* causes hybrid lethality through an epistatic interaction with the *mel* X, we switched the species origin of the X chromosome in hybrids while holding the rest of the genotype constant. Hybrid males that were heterozygous *l(3)95BCd*<sup>E53.1</sup>/*sim* and carried the *mel* X were inviable (*Df/Bal* ratio = 0.05, *n* = 101; Table 1); by contrast, hybrid males that had an identical genotype (including cytoplasm) and carried a *sim* X chromosome were viable (*Df/Bal* ratio = 1.29, *n* = 39). The *sim* allele of *l(3)95BCd* thus interacts with a gene(s) on the *mel* X chromosome to cause hybrid lethality.

We determined that the molecular lesions in *l(3)95BCd* mutant alleles affect the sequence CG10198, which represents the *Drosophila* homologue of a dicistronic gene encoding two functionally distinct nucleoporins (Nups), *Nup98-Nup96* (Fig. 2a). *Nup98* and *Nup96* are two of ~30 distinct Nups that function in the nuclear pore complex (NPC)<sup>16–19</sup>. NPCs are among the largest macromolecular complexes in eukaryotic cells and function as the sole site of cytonuclear trafficking of RNAs and proteins. The *Nup98* and *Nup96* proteins are both found on the nucleoplasmic and cytoplasmic sides of the NPC. However, *Nup98* is a mobile protein that shuttles on and off the NPC<sup>20–22</sup>, whereas *Nup96* is stably bound at the NPC, where it seems to have a structural role<sup>23</sup>. Both proteins function in RNA export<sup>21,24,25</sup>. A BLAST search of the entire *Drosophila* genome shows that *Nup98-Nup96* is a single-copy gene in *D. melanogaster*; Southern blot and sequence data indicate that the same is true in *D. simulans* (data not shown). As in other eukaryotes, *Drosophila Nup98-Nup96* is alternatively transcribed, giving rise to a minor 3.5-kilobase (kb) *Nup98* transcript and a major 7.3-kb *Nup98-Nup96* transcript. Northern blot data show that both transcripts are expressed at all developmental stages examined, from embryo to early adult (data not shown). The former encodes the *Nup98* protein, but the latter encodes a single, large precursor polypeptide that, in other eukaryotes, and thus presumably in flies, autoproteolytically cleaves itself between residues 1028(F) and 1029(S), yielding separate *Nup98* and *Nup96* proteins<sup>26–28</sup> (Fig. 2b). The 12 residues surrounding this cleavage site are nearly perfectly conserved in yeast, nematodes, rodents, humans and flies.

Complementation tests show that hybrid lethality is caused by the *D. simulans* (Ds) *Nup96* protein and suggest (but do not prove) that hybrid lethality involves its amino terminus. First, mutations that fail to complement hybrid lethality disrupt *D. melanogaster* (Dm) *Nup96* but leave Dm*Nup98* intact: two mutations, one with no (F1.13) and one with incomplete (339) complementation, produce truncated Dm*Nup96* and intact Dm*Nup98* (Table 1 and Fig. 2b); similarly, E53.1, which also uncovers hybrid lethality, produces no Dm*Nup96* and a nearly intact Dm*Nup98* (truncated just seven amino acids short of the autoproteolysis cleavage site; Table 1). These findings show that intact Dm*Nup98* is insufficient to rescue hybrid lethality and that disrupted Dm*Nup96* uncovers hybrid lethality. Second, the relevant region of Ds*Nup96* seems to be its N terminus. Hybrid lethality can be rescued by supplying hybrid males with a truncated form of Dm*Nup96* (mutation F1.15, which truncates Dm*Nup96* at residue 1142), indicating that the region between residues 1029 and 1142 is involved in hybrid lethality. Interestingly, this short region contains the only viability-essential portion of the *Nup98-Nup96* yeast homologue, *Nup145* (ref. 24). (That the longer Dm*Nup96*-F1.13 mutant allele does not rescue lethality is consistent with the possibility that its rescue is compromised by interfering higher-order structure and/or its intrinsic instability. Note, however, that as increasingly longer mutant alleles of Dm*Nup96* are used—for example, 339, *Df(3R)CA15*—hybrid viability is gradually restored; Table 1.) Together, these findings show that Ds*Nup96* causes hybrid lethality and they are at least consistent with the notion that lethality involves its N terminus.

### Adaptive evolution drove the substitutions in *Nup96*

We next studied the evolutionary history of *Nup98-Nup96*. Focusing first on divergence at *Nup98-Nup96* between *D. simulans* and *D. melanogaster*, we used a sliding window method to study rates of non-synonymous (amino-acid changing; *K<sub>a</sub>*) and synonymous (non-amino-acid changing; *K<sub>s</sub>*) divergence across the entire 5.9-kb coding sequence. *K<sub>a</sub>/K<sub>s</sub>* ratios greater than 1 represent extremely stringent, but definitive, evidence for positive natural selection<sup>29</sup>. We detect a dramatic peak in divergence with *K<sub>a</sub>/K<sub>s</sub>* > 1 (Fig. 2c). Although not significantly greater than 1 (the *K<sub>a</sub>/K<sub>s</sub>* > 1 standard is notoriously conservative), this striking divergence is at least suggestive of a history of positive natural selection in the vicinity of the relevant (incompatible) substitutions. To obtain a more detailed history and to perform a more powerful analysis, we surveyed DNA polymorphisms by sequencing the entire 2.8-kb *Nup96* gene from African populations of *D. melanogaster* and *D. simulans*. We studied alleles sampled from African populations because these are less likely to be confounded by demographic effects seen in recently founded cosmopolitan populations<sup>30</sup>. We analysed 15 alleles of *D. simulans* and 15 of *D. melanogaster* from isofemale lines collected in Zimbabwe. If the *Nup96* gene evolved under strict neutrality, only functionally unconstrained sites should vary within and between species. Thus, the neutral expectation is that the ratio of replacement (*R*) to silent (*S*) polymorphisms within species will be roughly

Table 1 Mutations disrupting *Nup96* fail to complement hybrid lethality

| Allele            | Mutation cDNA position | Nucleotide change | Amino-acid position | Amino-acid change | Affected protein | Hybrid males recovered |                | <i>Df/Bal</i> ratio | Lethal in hybrids? |
|-------------------|------------------------|-------------------|---------------------|-------------------|------------------|------------------------|----------------|---------------------|--------------------|
|                   |                        |                   |                     |                   |                  | <i>Df/sim</i>          | <i>Bal/sim</i> |                     |                    |
| <b>E53.1</b>      | <b>3063</b>            | <b>TGG to TGA</b> | <b>1021</b>         | <b>W to STOP</b>  | <b>Nup98</b>     | <b>5</b>               | <b>96</b>      | <b>0.052</b>        | <b>Yes</b>         |
| F8.5              | 3115                   | GAG to CAG        | 1039                | E to Q            | Nup96            | 83                     | 47             | 1.766               | No                 |
| F1.15             | 3425                   | TTG to TAG        | 1142                | L to STOP         | Nup96            | 59                     | 35             | 1.686               | No                 |
| <b>F1.13</b>      | <b>4600</b>            | <b>CAG to TAG</b> | <b>1534</b>         | <b>Q to STOP</b>  | <b>Nup96</b>     | <b>2</b>               | <b>35</b>      | <b>0.057</b>        | <b>Yes</b>         |
| <b>339</b>        | <b>5176</b>            | <b>CAG to TAG</b> | <b>1726</b>         | <b>Q to STOP</b>  | <b>Nup96</b>     | <b>19</b>              | <b>62</b>      | <b>0.306</b>        | <b>Partial</b>     |
| C14.7             | —                      | —                 | —                   | —                 | —                | 90                     | 26             | 3.462               | No                 |
| <i>Df(3R)CA15</i> | ≥5074                  | Deletion          | ≥1693               | Deletion          | Nup96            | 178                    | 162            | 1.099               | No                 |

Mutant alleles of the *D. melanogaster Nup98-Nup96* gene were tested for their ability to uncover hybrid lethality when heterozygous with the *D. simulans* wild-type allele. *D. melanogaster* females heterozygous for a mutation over a dominantly marked balancer were crossed to *D. simulans Lhr* males. The ratio of mutation- to balancer-inheriting hybrid sons was scored (see Fig. 1). Mutations that uncover hybrid lethality are in bold font.



Table 2 Adaptive evolution caused *Nup96* divergence between species

|   |                                                    | Polymorphic |     |           | Divergent |     |           | G-value | P-value |
|---|----------------------------------------------------|-------------|-----|-----------|-----------|-----|-----------|---------|---------|
|   |                                                    | R           | S   | R/S ratio | R         | S   | R/S ratio |         |         |
| 1 | <i>D. melanogaster</i> versus <i>D. simulans</i> * | 27          | 108 | 0.250     | 27        | 34  | 0.794     | 11.888  | 0.00056 |
| 2 | <i>D. melanogaster</i> versus <i>D. yakuba</i>     | 5           | 43  | 0.116     | 69        | 152 | 0.454     | 9.984   | 0.00158 |
| 3 | <i>D. simulans</i> versus <i>D. yakuba</i>         | 21          | 68  | 0.309     | 60        | 139 | 0.432     | 1.334   | 0.24803 |
| 4 | <i>D. melanogaster</i> versus <i>D. mauritiana</i> | 5           | 43  | 0.116     | 32        | 51  | 0.627     | 13.202  | 0.00028 |
| 5 | <i>D. simulans</i> versus <i>D. mauritiana</i>     | 22          | 69  | 0.319     | 3         | 13  | 0.231     | 0.233   | 0.62911 |
| 6 | <i>D. melanogaster</i> lineage†                    | 5           | 43  | 0.116     | 16        | 21  | 0.762     | 12.351  | 0.0012  |
| 7 | <i>D. simulans</i> lineage                         | 22          | 69  | 0.319     | 10        | 8   | 1.250     | 6.567   | 0.0104  |

\* Pooled polymorphism data.

† Substitutions that could not be unambiguously assigned to either lineage were excluded.

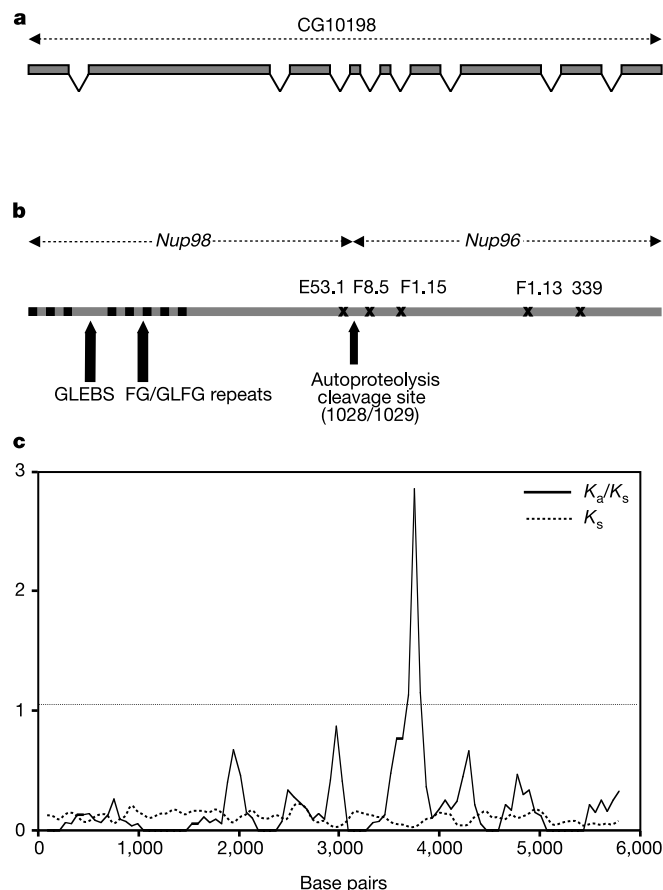
equal to the *R/S* ratio of fixed differences between species<sup>31</sup>. Our data, however, strongly reject neutrality for *Nup96*: there is a highly significant excess of replacement changes fixed between species, indicating a history of adaptive evolution (Table 2, line 1).

After sequencing single alleles from two related species, *D. yakuba* and *D. mauritiana*, we mapped substitutions onto the known phylogeny of the *D. melanogaster* group<sup>32</sup> and, in particular, onto the *D. melanogaster* and *D. simulans* lineages (Fig. 3). Comparing the *R/S* ratio of substitutions that occurred in the lineages leading to *D. melanogaster* and *D. simulans* to the *R/S* ratio of polymorphisms within each species, respectively, reveals that adaptive evolution drove an excess of replacement substitutions in both species' histories (Table 2, lines 6, 7). Two lines of evidence suggest that

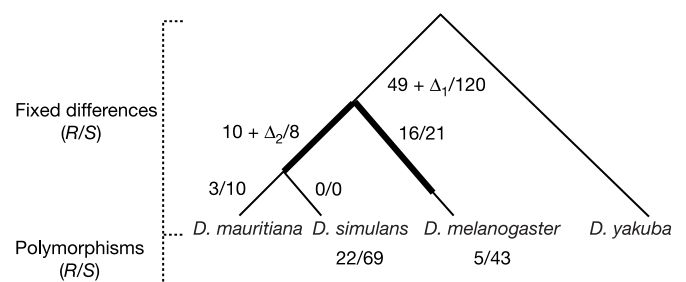
these bouts of adaptation occurred in the fairly distant past. First, all substitutions in the *D. simulans* lineage occurred before the split with *D. mauritiana* (that is, >0.26 Myr ago; Fig. 3). Second, there is no evidence of a recent selective sweep in either species. Such an event would leave a signature of reduced nucleotide diversity in the region<sup>33,34</sup> and characteristic patterns in the distribution of allele frequencies—namely, an excess of rare variants<sup>35</sup> and an excess of high-frequency derived variants<sup>36</sup>. However, mean synonymous nucleotide diversities in the coding regions of *Nup96* are typical, if not slightly high, for autosomal loci sampled from African populations<sup>37</sup>:  $\pi = 0.0193$  in *D. melanogaster* and  $\pi = 0.0285$  in *D. simulans* ( $\pi = 4N_e\mu$ , where  $N_e$  is effective population size and  $\mu$  is the per-site mutation rate). Moreover, neither species' frequency spectrum deviates significantly from neutral equilibrium expectations using the tests of Tajima<sup>38</sup> and of Fay and Wu<sup>36</sup> ( $P \geq 0.335$  in all tests). Therefore, the adaptive substitutions fixed at *Nup96* do not seem to be recent events in either species.

## Conclusions

Our molecular and population genetic analyses of *Nup96* allow us to address a number of issues in speciation genetics<sup>1,2</sup>. First, a single gene can explain the hybrid lethality of a small chromosomal region identified in our deficiency screen. This finding contrasts with those from studies of hybrid male sterility that have found that several tightly linked incompatibility factors seem to be required to cause complete hybrid sterility<sup>39,40</sup>. Second, *Nup96* is a viability-essential gene within species, performing a fundamental cell-biological function. Thus, this instance of a hybrid incompatibility involves an 'ordinary' gene, not a selfishly propagating genomic parasite (for example, repetitive DNA or transposon) as previous workers have speculated<sup>41</sup>. Third, *Nup96* evolved by positive natural selection in both species' lineages. As an incidental by-product of this adaptation, the *D. simulans* *Nup96* protein is no longer compatible with an (unknown) interacting factor(s) encoded by the *D. melanogaster* X chromosome. Fourth, our functional and population genetic data



**Figure 2** *Nup98-Nup96* structure and evolution. **a**, *Nup98-Nup96* (CG10198) gene structure. **b**, *Nup98* and *Nup96* protein structure with several conserved motifs indicated: GLE-binding site in *Nup98*; FG/GLFG repeats in *Nup98*; and autoproteolysis cleavage site. Locations of molecular lesions in *l(3)95BCd* mutant alleles are indicated by x. **c**, Sliding window analysis of the ratio of non-synonymous ( $K_a$ ) and synonymous ( $K_s$ ) substitution rates. Window size, 180 base pairs (bp).



**Figure 3** Evolutionary history of *Nup96* with ratios of replacement to silent substitutions mapped onto the known phylogeny of the *D. melanogaster* group species. *R/S* ratios of fixed differences are shown on the branches of the phylogeny; polymorphisms in *D. simulans* and *D. melanogaster* are shown at the tips of phylogeny. Bold branches indicate those in which *Nup96* experienced adaptive evolution (for statistics, see Table 2, lines 6, 7).  $\Delta_1$ , a 12-bp, in-frame indel;  $\Delta_2$ , a 3-bp, in-frame insertion.

implicate structural divergence in the DsNup96 protein. This case of hybrid inviability does not therefore seem to involve the divergence of *cis*-regulatory sequences<sup>42</sup>. Last, *Nup96* is a single-copy gene. Its role in hybrid inviability could not therefore have evolved in accordance with a recent theory that suggests that hybrid incompatibilities arise as by-products of divergence among duplicate genes<sup>43</sup>.

Several questions about our hybrid inviability gene remain. First, as *Nup96* encodes a stable constituent of NPCs, we speculate that nuclear transport is disrupted in inviable hybrid males that are hemizygous for *DsNup96* and that carry an incompatible *D. melanogaster* X chromosome, but not in (viable) hybrid males that carry the co-adapted *D. simulans* X chromosome. It will be of great interest to determine whether the (at present unknown) interacting factor encoded by the X chromosome shows evidence of correlated adaptive evolution. Second, we know nothing about the specific selective forces that caused the adaptive evolution of *Nup96* in each species' history. Indeed, this adaptive evolution is surprising given that NPCs are otherwise remarkably conserved in architecture<sup>44</sup>, in the number of nucleoporins<sup>18,19</sup> and in the functions of homologous nucleoporins<sup>16,17</sup> among eukaryotes.

Whether the conclusions drawn from *Nup96* and from the few other hybrid incompatibility genes studied so far will prove general remains to be seen. But the current analysis shows that our screen for recessive hybrid lethals is effective and should yield information on the identities and evolutionary histories of many other hybrid incompatibility genes. □

## Methods

### Crosses

All crosses were performed at 24 °C and flies were reared on standard cornmeal–yeast–agar medium. Species crosses were made by mass mating 15–20 *D. melanogaster* females to 15–25 *D. simulans* males. *Nup96* was mapped by crossing *D. melanogaster* females heterozygous for a deficiency or a recessive lethal mutation over a dominantly marked balancer (for example, TM3, *Sb*) to *D. simulans* males carrying *Lhr* (Fig. 1). *Lhr* rescues the normally dead hybrid males from this species cross<sup>15</sup>, exposing them to other (recessive) hybrid incompatibilities unmasked by the deficiencies or mutations used<sup>11</sup> (Fig. 1). We scored the number of hybrid males inheriting the deficiency or loss-of-function mutation and those inheriting the balancer. Deficiencies were considered lethal when the ratio of deficiency-carrying (or mutation-carrying) hybrid males to balancer-carrying hybrid males was  $\leq 0.10$ . Deficiency- and balancer-inheriting hybrid females from these crosses were always viable.

We established the genetic breakpoints of *Df(3R)CA15* and *Df(3R)mbc-BG1* by complementation tests within *D. melanogaster*. These deficiencies were produced concurrently with the others described in refs 45, 46. *Df(3R)CA15* fails to complement mutations affecting *l(3)04684* through *l(3)95BCd* (Fig. 1b). We focused on CG10198 as the molecular candidate for *l(3)95BCd* because Southern analysis shows that the distal breakpoint of *Df(3R)CA15* truncates the 3'-end of CG10198 (negative strand) but does not affect the distal-adjacent gene *mbc* (positive strand), thereby splitting these two genes. Sequencing of CG10198 from *l(3)95BCd* mutant lines confirmed the presence of molecular lesions (Table 1).

To test the epistatic basis of the *sim l(3)95BCd* allele's hybrid lethality, the incompatible *mel* X chromosome was replaced with a compatible *sim* X chromosome in hybrid males. This was accomplished using a *mel* attached-X chromosome (two X chromosomes fused to a single centromere) to enforce paternal inheritance of the *sim* X chromosome<sup>11</sup>. (In *Drosophila*, sex is determined by the ratio of the number of Xs to that of each autosome.) Briefly, we crossed *D. melanogaster* *C(1)M4, y<sup>2</sup>; l(3)95BCd<sup>E53.1</sup>/TM3, Tb Sb* females to *D. simulans* *Lhr* males. Offspring inheriting the *mel* attached-X chromosome develop as females homozygous for the *D. melanogaster* X, whereas those inheriting their only X from their *sim* father develop as hybrid males hemizygous for the *D. simulans* X chromosome. Note that hybrid sons from this cross inherit their cytoplasm (and all associated maternal factors) from their *D. melanogaster* mothers, so that these hybrid males are genotypically identical to those from the original cross, except at X-linked loci. Hybrids from this cross were scored as above.

### Sequencing

Primers designed from the annotated *D. melanogaster* genome sequence, CG10198, were used for polymerase chain reaction (PCR) amplification of genomic DNA, followed by direct sequencing of both strands of the PCR products. Sequencing was done using ABI prism BigDye Chemistry (Perkin Elmer) on an automated ABI sequencer. Lethal mutant alleles of *l(3)95BCd* were balanced over TM3-GFP (green fluorescent protein) chromosomes within *D. melanogaster* and *Nup98-Nup96* was sequenced from genomic DNA extracted from homozygous non-GFP embryos. Wild-type alleles were sequenced from genomic DNA extracted from 15 isofemale lines of *D. melanogaster* collected in Zimbabwe; 15 isofemale lines of *D. simulans* collected in Zimbabwe; *D. simulans* *Lhr*; *D. mauritiana* 0214-6; and *D. yakuba* Tai 15.

### Sequence analysis

All sequences were edited using Sequencher version 3.0 and then manually aligned in SeAl version 1.0. Sliding window estimation of  $K_a/K_s$  was performed using *K-estimator*<sup>47</sup>. Fay and Wu's<sup>36</sup> *H*-test was performed using the online program available at <http://crimp.lbl.gov/hstest.html>; all other population genetic analyses were performed using the DnaSP program<sup>48</sup>.

Received 24 December 2002; accepted 10 March 2003; doi:10.1038/nature01679.

- Coyne, J. Genetics and speciation. *Nature* **355**, 511–515 (1992).
- Orr, H. A. & Presgraves, D. C. Speciation by postzygotic isolation: forces, genes and molecules. *Bioessays* **22**, 1085–1094 (2000).
- Wittbrodt, J. et al. Novel putative receptor tyrosine kinase encoded by the melanoma-inducing *Tu* locus in *Xiphophorus*. *Nature* **341**, 415–421 (1989).
- Ting, C.-T., Tsaur, S.-C., Wu, M.-L. & Wu, C.-I. A rapidly evolving homeobox at the site of a hybrid sterility gene. *Science* **282**, 1501–1504 (1998).
- Barbash, D. A., Siino, D. F., Tarone, A. M. & Roote, J. A rapidly evolving Myb-related protein causes species isolation in *Drosophila*. *Proc. Natl Acad. Sci. USA* **100**, 5302–5307 (2003).
- Sturtevant, A. H. The genetics of *Drosophila simulans*. *Yb. Carnegie Inst. Wash.* **399**, 1–62 (1929).
- Sturtevant, A. H. Genetic studies on *Drosophila simulans*. I. Introduction. Hybrids with *Drosophila melanogaster*. *Genetics* **5**, 488–500 (1920).
- Lachaise, D., David, J. R., Lemeunier, F., Tsacas, L. & Ashburner, M. The reproductive relationships of *Drosophila sechellia* with *D. mauritiana*, *D. simulans*, and *D. melanogaster* from the Afrotropical region. *Evolution* **1986**, 262–271 (1986).
- Davis, A. W. et al. Rescue of hybrid sterility in crosses between *D. melanogaster* and *D. simulans*. *Nature* **380**, 157–159 (1996).
- Sawamura, K., Davis, A. W. & Wu, C.-I. Genetic analysis of speciation by means of introgression into *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **97**, 2652–2655 (2000).
- Presgraves, D. C. A fine-scale genetic analysis of hybrid incompatibilities in *Drosophila*. *Genetics* **163**, 955–972 (2003).
- Dobzhansky, T. *Genetics and the Origin of Species* (Columbia Univ. Press, New York, 1937).
- Muller, H. J. & Pontecorvo, G. Recessive genes causing interspecific sterility and other disharmonies between *Drosophila melanogaster* and *simulans*. *Genetics* **27**, 157 (1942).
- Turelli, M. & Orr, H. A. Dominance, epistasis and the genetics of postzygotic isolation. *Genetics* **154**, 1663–1679 (2000).
- Watanabe, T. K. A gene that rescues the lethal hybrids between *Drosophila melanogaster* and *D. simulans*. *Jpn. J. Genet.* **54**, 325–331 (1979).
- Ryan, K. J. & Wente, S. R. The nuclear pore complex: a protein machine bridging the nucleus and the cytoplasm. *Curr. Opin. Cell Biol.* **12**, 361–371 (2000).
- Vasu, S. K. & Forbes, D. J. Nuclear pores and nuclear assembly. *Curr. Opin. Cell Biol.* **13**, 363–375 (2001).
- Rout, M. P. et al. The yeast nuclear pore complex: composition, architecture and transport mechanism. *J. Cell Biol.* **148**, 635–651 (2000).
- Cronshaw, J. M., Krutchinsky, A. N., Zhang, W., Chait, B. T. & Matunis, M. J. Proteomic analysis of the mammalian nuclear pore complex. *J. Cell Biol.* **158**, 915–927 (2002).
- Fontoura, B. M. A., Dales, S., Blobel, G. & Zhong, H. The nucleoporin Nup98 associates with the intranuclear filamentous protein network of TPR. *Proc. Natl Acad. Sci. USA* **98**, 3208–3213 (2001).
- Griffis, E. R., Altan, N., Lippincott-Schwartz, J. & Powers, M. A. Nup98 is a mobile nucleoporin with transcription dependent dynamics. *Mol. Biol. Cell.* **13**, 1282–1297 (2002).
- Griffis, E. R., Xu, S. & Powers, M. A. Nup98 localizes to both nuclear and cytoplasmic sides of the nuclear pore and binds to two distinct nucleoporin subcomplexes. *Mol. Biol. Cell.* **14**, 600–610 (2003).
- Belgareh, N. et al. An evolutionarily conserved NPC subcomplex, which redistributes in part to kinetochores in mammalian cells. *J. Cell Biol.* **154**, 1147–1160 (2001).
- Emtage, J. L. T., Buccini, M., Watkins, J. L. & Wente, S. R. Defining the essential functional regions of the nucleoporin Nup145p. *J. Cell Sci.* **110**, 911–925 (1997).
- Powers, M. A., Forbes, D. J., Dahlberg, D. J. & Lund, E. The vertebrate GLFG nucleoporin, Nup98, is an essential component of multiple RNA export pathways. *J. Cell Biol.* **136**, 241–250 (1997).
- Fontoura, B. M. A., Blobel, G. & Matunis, M. J. A conserved biogenesis pathway for nucleoporins: proteolytic processing of a 186-Kilodalton precursor generates Nup98 and the novel nucleoporin, Nup96. *J. Cell Biol.* **144**, 1097–1112 (1999).
- Rosenblum, J. S. & Blobel, G. Autophosphorylation in nucleoporin biogenesis. *Proc. Natl Acad. Sci. USA* **96**, 11370–11375 (1999).
- Teixeira, M. T., Fabre, E. & Dujon, B. Self-catalyzed cleavage of the yeast nucleoporin Nup145 precursor. *J. Biol. Chem.* **274**, 32439–32444 (1999).
- Kimura, M. Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. *Nature* **267**, 275–276 (1977).
- Wall, J. D., Andolfatto, P. & Przeworski, M. Testing models of selection and demography in *Drosophila simulans*. *Genetics* **162**, 203–216 (2002).
- McDonald, J. H. & Kreitman, M. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* **351**, 652–654 (1991).
- Hey, J. & Kliman, R. M. Population genetics and phylogenetics of DNA sequence variation at multiple loci within the *Drosophila melanogaster* species complex. *Mol. Biol. Evol.* **10**, 804–822 (1993).
- Maynard Smith, J. & Haigh, J. The hitch-hiking effect of a favourable gene. *Genet. Res.* **23**, 23–35 (1974).
- Kaplan, N. L., Hudson, R. R. & Langley, C. H. The hitch-hiking effect revisited. *Genetics* **123**, 887–899 (1989).
- Braverman, J. M., Hudson, R. R., Kaplan, N. L., Langley, C. H. & Stephan, W. The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. *Genetics* **140**, 783–796 (1995).
- Fay, J. C. & Wu, C.-I. Hitchhiking under positive Darwinian selection. *Genetics* **155**, 1405–1413 (2000).
- Andolfatto, P. Contrasting patterns of X-linked and autosomal nucleotide variation in *Drosophila melanogaster* and *Drosophila simulans*. *Mol. Biol. Evol.* **18**, 279–290 (2001).
- Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595 (1989).

39. Naveira, H. F. & Maside, X. R. in *Endless Forms* (eds Howard, D. J. & Berlocher, S. H.) 330–338 (Oxford Univ. Press, Oxford, 1998).
40. Wu, C.-I. & Hollocher, H. in *Endless Forms* (eds Howard, D. J. & Berlocher, S. H.) 339–351 (Oxford Univ. Press, Oxford, 1998).
41. Rose, M. & Doolittle, W. F. Molecular biological mechanisms of speciation. *Science* **220**, 157–162 (1983).
42. Wilson, A. C. in *Molecular Evolution* (ed. Ayala, F. J.) 225–234 (Sinauer, Sunderland, Massachusetts, 1976).
43. Lynch, M. & Force, A. G. The origin of interspecific genomic incompatibility via gene duplication. *Am. Nat.* **156**, 590–605 (2000).
44. Yang, Q., Rout, M. P. & Akey, C. W. Three-dimensional architecture of the isolated yeast nuclear pore complex: functional and evolutionary implications. *Mol. Cell* **1**, 223–234 (1998).
45. Erickson, M. R. S., Galletta, B. J. & Abmayr, S. M. *Drosophila myoblast city* encodes a conserved protein that is essential for myoblast fusion, dorsal closure and cytoskeletal organization. *J. Cell Biol.* **138**, 589–603 (1997).
46. Keller, C. A., Grill, M. A. & Abmayr, S. M. A role for nautilus in the differentiation of muscle precursors. *Dev. Biol.* **202**, 157–171 (1998).
47. Comeron, J. M. K-estimator: calculation of the number of nucleotide substitutions per site and the confidence intervals. *Bioinformatics* **15**, 763–764 (1999).
48. Rozas, J. & Rozas, R. DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* **15**, 174–175 (1999).

**Acknowledgements** We thank J. Coyne for providing the *D. yakuba* stock, and C. Aquadro for the Zimbabwe strains of *D. melanogaster* and *D. simulans*. We acknowledge the contribution of M. Erickson in preliminary molecular analysis of the nucleoporin locus. We thank A. Betancourt, A. Clark, J. Coyne, W. Stephan and J. Werren for helpful discussion and comments. This work was supported by funds from a Caspari Fellowship, a Messermith Fellowship, and a Dissertation Improvement Grant from the National Science Foundation to D.C.P.; from the National Institutes of Health and National Science Foundation to S.M.A.; and from the National Institutes of Health to H.A.O.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to D.C.P. (dvnnp@mail.rochester.edu). Sequences have been deposited in GenBank under accession numbers AY250768–AY250800.



# Functional proteomic identification of DNA replication proteins by induced proteolysis *in vivo*

Masato Kanemaki\*, Alberto Sanchez-Diaz\*, Agnieszka Gambus & Karim Labib

\* Cancer Research UK, Paterson Institute for Cancer Research, Christie Hospital NHS Trust, Wilmslow Road, Manchester, M20 4BX, UK

\*These authors contributed equally to this work

Evolutionarily diverse eukaryotic cells share many conserved proteins of unknown function. Some are essential for cell viability<sup>1,2</sup>, emphasising their importance for fundamental processes of cell biology but complicating their analysis. We have developed an approach to the large-scale characterization of such proteins, based on conditional and rapid degradation of the target protein *in vivo*, so that the immediate consequences of bulk protein depletion can be examined<sup>3</sup>. Budding yeast strains have been constructed in which essential proteins of unknown function have been fused to a 'heat-inducible-degron' cassette that targets the protein for proteolysis at 37 °C (ref. 4). By screening the collection for defects in cell-cycle progression, here we identify three DNA replication factors that interact with each other and that have uncharacterized homologues in human cells. We have used the degron strains to show that these proteins are required for the establishment and normal progression of DNA replication forks. The degron collection could also be used to identify other, essential, proteins with roles in many other processes of eukaryotic cell biology.

Analysing the role of eukaryotic proteins that are essential for cell viability requires a method by which protein function can be inactivated conditionally. Until now, the only methods suitable for characterizing large numbers of such proteins involve indirect approaches to protein inactivation, either by replacing the promoter of the encoding gene with another that is repressible<sup>5–7</sup>, or else by RNA interference<sup>1</sup>. In these cases the messenger RNA can rapidly be depleted, but decay of the protein is often much slower, so that the phenotype may be complicated by secondary responses to the initial consequences of partially inactivating the protein.

Recently we described a simple method by which proteins of the budding yeast *Saccharomyces cerevisiae* can be depleted rapidly in a manner that is direct and conditional<sup>3</sup>. The chromosomal locus of a gene is modified so that the amino-terminus of the encoded protein is fused to a 'heat-inducible-degron' cassette that targets the fusion protein for degradation at 37 °C (ref. 4). The degron is recognized by the Ubr1 protein that is associated with a ubiquitin-conjugating enzyme, and proteolysis is stimulated at higher temperatures, probably because unfolding of the degron cassette exposes lysine residues that are the sites of ubiquitylation<sup>4</sup>. We showed previously that essential proteins fused to the degron are functional at 37 °C in the absence of Ubr1, but are degraded efficiently at 37 °C when the Ubr1 protein is subsequently expressed to high levels<sup>3</sup>.

To compare direct inactivation of a protein using the degron method with depletion of the mRNA by repressing transcription, we inactivated the *S. cerevisiae* DNA replication protein Mcm4 using the two approaches. When inactivation was assayed by the ability of cells to form colonies under restrictive conditions, both methods appear equally efficient (Fig. 1a). However, more careful analysis of the phenotype in liquid culture experiments revealed important differences. Repressing transcription of the *MCM4* gene produced a rapid decrease in the level of *MCM4* mRNA, but depletion of Mcm4 protein was more gradual (Fig. 1b). This caused cells to become arrested with a DNA content of 2C, typical of defects in G2 phase or mitosis, probably because subtle problems in chromosome replication occur as Mcm4 is gradually depleted, activating a checkpoint response that blocks completion of the cell cycle. In contrast, Mcm4 protein was rapidly depleted in the degron strain, despite persistence of the mRNA, producing a marked and obvious defect in

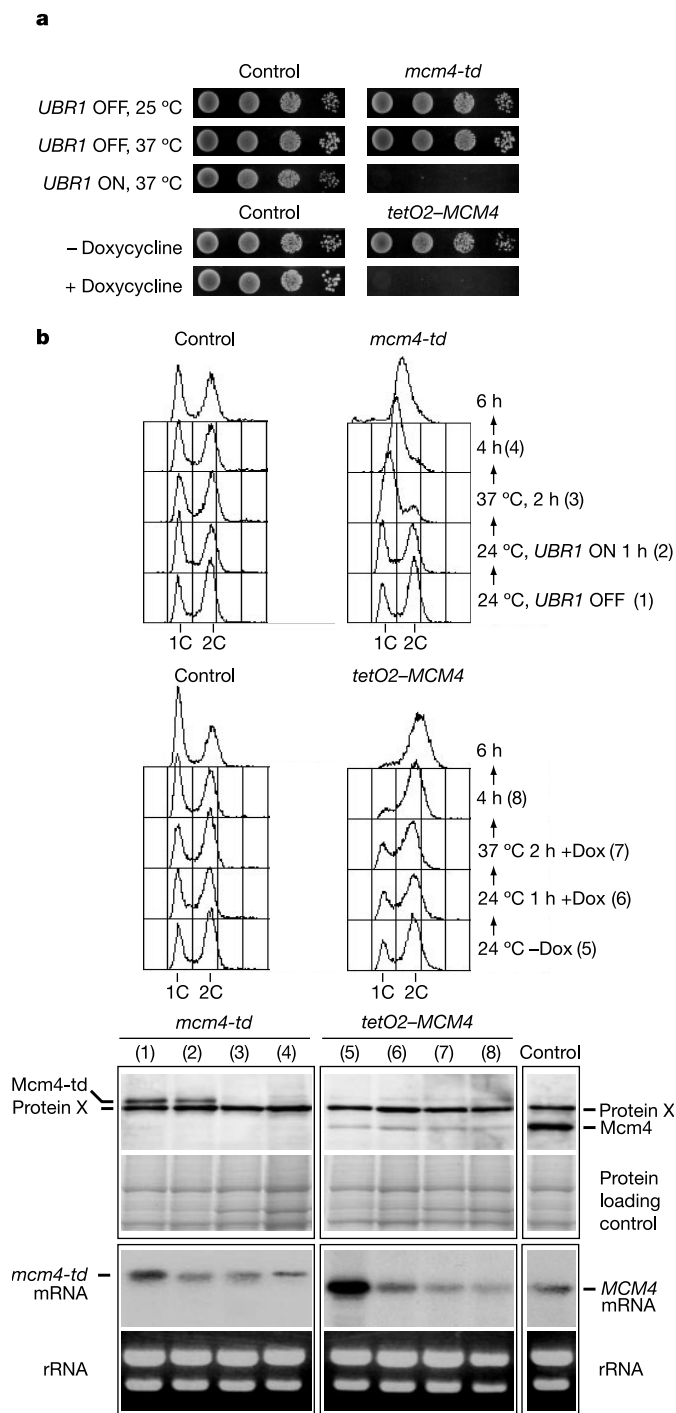
chromosome replication (Fig. 1b). This qualitative difference in phenotype demonstrates the importance of rapidly depleting proteins essential for cell viability, rather than their corresponding mRNAs, in order to assign a biological function.

We have made yeast strains in which the degron tag has been added to more than half of the essential *S. cerevisiae* proteins of unknown function, using a one-step polymerase chain reaction (PCR) method (see Methods and Supplementary Table 1 for a complete list of the proteins). Most of these proteins have homologues of unknown function in other eukaryotes, and their functions cannot be predicted from the amino acid sequence. We were able to make viable yeast strains carrying degron fusions for 94% of the essential proteins tested, and about 60% of these are specifically unable to form colonies at 37 °C in the presence of Ubr1 (the full data set is described in Supplementary Table 1 and Supplementary Fig. 1).

The resulting collection can be screened for defects in many fundamental processes of cell biology. Initially, we have screened the collection for strains defective in progression through the cell cycle at 37 °C. For each of the temperature-sensitive strains, we grew an asynchronous culture at 24 °C and then expressed Ubr1 before raising the temperature to 37 °C and examining cellular and nuclear morphology together with DNA content. In addition, we screened the degron strains that are viable at 37 °C (suggesting insufficient inactivation of the protein; see Supplementary Fig. 2) for ones where cell size was markedly increased at the high temperature, indicating a 'leaky' defect in cell-cycle progression.

In a control yeast strain, growth was unaffected after shifting to 37 °C after a brief period of adaptation (Fig. 2). Inactivating most of the essential proteins of unknown function blocked cell growth in a manner that does not correlate with position in the cell cycle. In contrast, however, we identified three new DNA replication factors that we call Cdc101, Cdc102 and Cdc105 (proteins with other roles in the cell cycle will be described elsewhere). For two of these—Cdc102/Yjl072c and Cdc105/Ydr489w—the corresponding degron strains are specifically unable to grow at 37 °C in the presence of high levels of Ubr1 (Fig. 2a) and accumulate as large-budded cells with largely unreplicated DNA (Fig. 2b; see also Supplementary Fig. 3). The Cdc101/Ydr013w degron strain (*cdc101-td*, where *td* denotes a

temperature-sensitive degon) is leaky, as it is able to grow at 37 °C even in the presence of Ubr1 (Fig. 2a); however, we observed a clear, although transient, defect in chromosome replication, and once again there is a marked accumulation of large-budded cells with a single nucleus (Fig. 2b; see also Supplementary Fig. 3).



**Figure 1** Direct inactivation of essential proteins by fusion to a heat-inducible degon. **a**, Serial dilutions were made of an *MCM4* degon strain (*mcm4-td*) or a strain in which transcription of *MCM4* can be repressed (*tetO2-MCM4*), together with appropriate control strains. **b**, The DNA content of *mcm4-td*, *tetO2-MCM4* and control strains was compared after transferring cells from permissive to restrictive conditions (identical except 20  $\mu\text{g ml}^{-1}$  doxycycline was added to the *tetO2-MCM4* strain and its control). The level of *MCM4* mRNA and Mcm4 protein was determined during the same experiment (the protein loading control shows the equivalent portion of a gel stained with Coomassie blue).

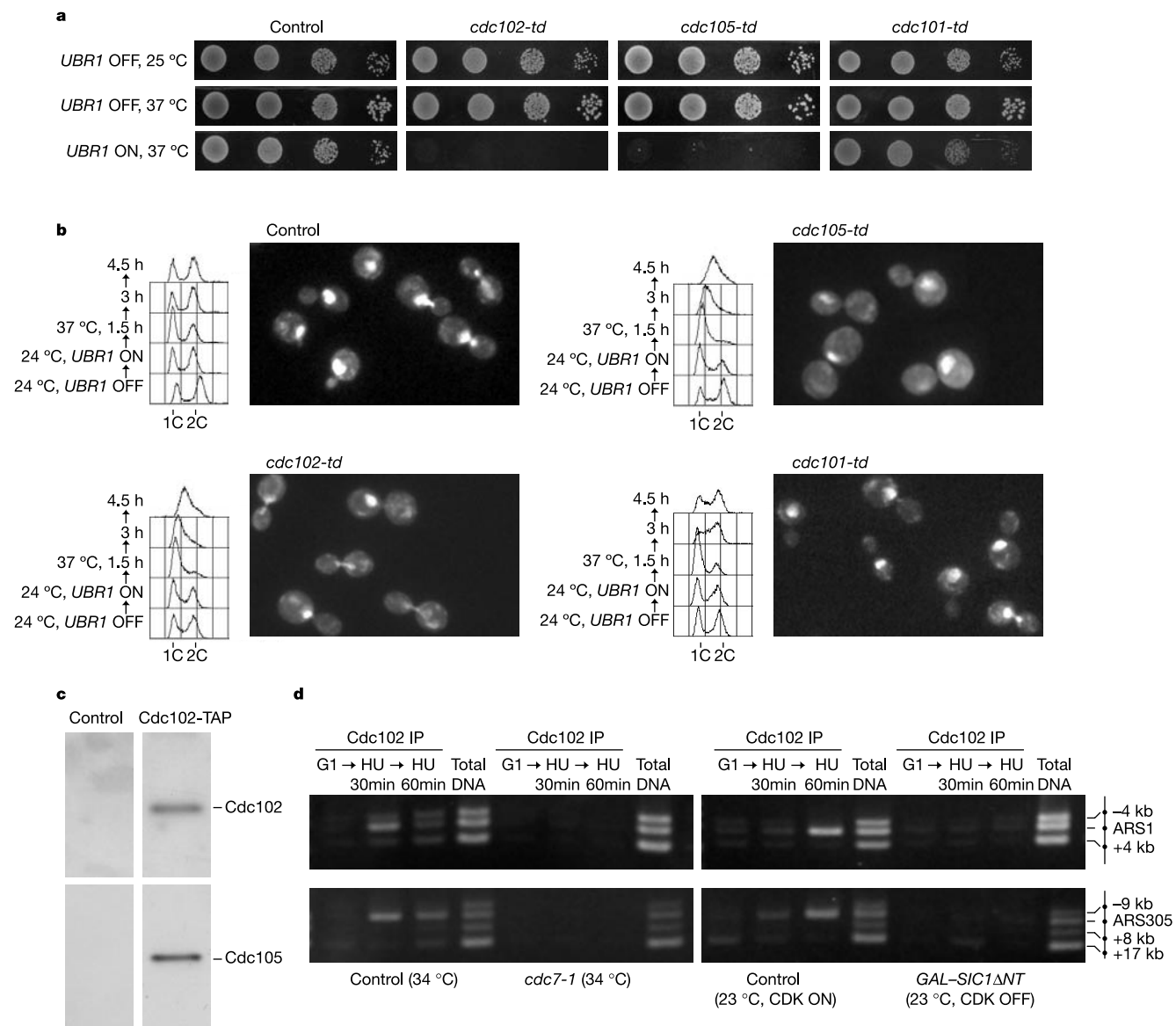
The Cdc101, Cdc102 and Cdc105 proteins are unrelated to each other, but all three have homologues of unknown function in other eukaryotes, including humans. Their function cannot be predicted from their amino acid sequence, but all three proteins interact with each other in genome-wide two-hybrid screens<sup>8,9</sup>. We purified Cdc102 from yeast extracts and used antibodies to Cdc102 and Cdc105 to show that these proteins do indeed form part of a new complex required for chromosome replication (Fig. 2c). We used mass spectrometry to identify two other components of the complex: Cdc101 and Yol146w (data not shown). Yol146w is essential and of unknown function, but the corresponding degon strain is viable at 37 °C (Supplementary Fig. 1). The same four proteins have recently been shown to form a complex required for an early step of chromosome replication, both in *S. cerevisiae* and in extracts of *Xenopus laevis* eggs<sup>10,11</sup>.

To examine the localization of Cdc102, we made yeast strains in which green fluorescent protein (GFP) or nine copies of the c-Myc epitope was added to the carboxy terminus of Cdc102 so that the mRNA expressing each fusion protein was still expressed from the *CDC102* promoter as the only version in the cell. Cdc102-GFP showed nuclear localization throughout the cell cycle (not shown), consistent with its role in chromosome replication. We then used chromatin immunoprecipitation (ChIP) experiments to examine the association of Cdc102-9Myc with replicating DNA. We released cells from G1 arrest in the presence of hydroxyurea, which inhibits ribonucleotide reductase and so reduces cellular dNTP pools. Under such conditions, initiation occurs at early origins but DNA replication forks then stall within a few kilobases<sup>12,13</sup>. Cdc102-9Myc associated specifically with early origins in hydroxyurea (Fig. 2d), with similar kinetics to the establishment of DNA replication forks<sup>12</sup>. Loading of Cdc102-9Myc required activation of both the Cdc7 kinase and the cyclin-dependent kinase Cdc28 (Fig. 2d), which are essential for the initiation of chromosome replication<sup>14-16</sup>. These experiments indicate that the complex containing Cdc102 and Cdc105 has a direct role in chromosome replication downstream of the two kinases that together trigger initiation.

We then examined the role of Cdc102 and Cdc105 in the establishment and progression of DNA replication forks *in vivo* at individual replicons. The Cdc102 and Cdc105 degon strains (*cdc102-td* and *cdc105-td*, respectively), together with a control strain, were grown for seven generations in medium containing 'heavy' isotopes of carbon and nitrogen, so that the chromosomal DNA was fully substituted (see Methods). Cells were then switched into 'light' medium and synchronized in the G1 phase of the cell cycle using mating pheromone. This procedure allowed us subsequently to distinguish unreplicated heavy-heavy DNA from replicated heavy-light DNA in CsCl gradients, by virtue of their differing densities<sup>17</sup>. In combination with 'slot blots' using probes for specific genomic sequences it is thus possible to follow the establishment and progression of individual DNA replication forks. Expression of Ubr1 was induced in the G1-arrested cells and the temperature was raised to 37 °C to inactivate Cdc102 or Cdc105. We then released cells from G1 arrest and examined their ability to proceed with chromosome replication.

In contrast to the control strain, which rapidly completed DNA replication within 60 min, very little replication was observed by flow cytometry after inactivation of Cdc102 or Cdc105, even after 90 min (Fig. 3a), although cells formed buds with normal kinetics (not shown). We assayed the replication of two of the earliest replicons in the *S. cerevisiae* genome: between the origin ARS306 and the left end of chromosome III, and between ARS607 and the right end of chromosome VI. Replication of each origin was found to be profoundly defective (Fig. 3b, probes 1 and 3), indicating that Cdc102 and Cdc105 are required for the establishment of DNA replication forks at these sites. Consistent with this defect, replication of the end of each replicon was similarly defective (Fig. 3b, probes 2 and 4).

To determine whether Cdc102 and Cdc105 are important for the



**Figure 2** Three new DNA replication proteins identified by functional proteomics. **a**, Serial dilutions of control or degon strains were grown under the indicated conditions and examined after 48 h. **b**, Liquid culture experiments for the same strains. DNA content was measured by flow cytometry throughout the experiment (left panels). Cells from the 3-h time point were stained with the DNA-binding dye 4,6-diamidino-2-phenylindole (DAPI) and nuclear morphology was examined by fluorescence microscopy (right panels).

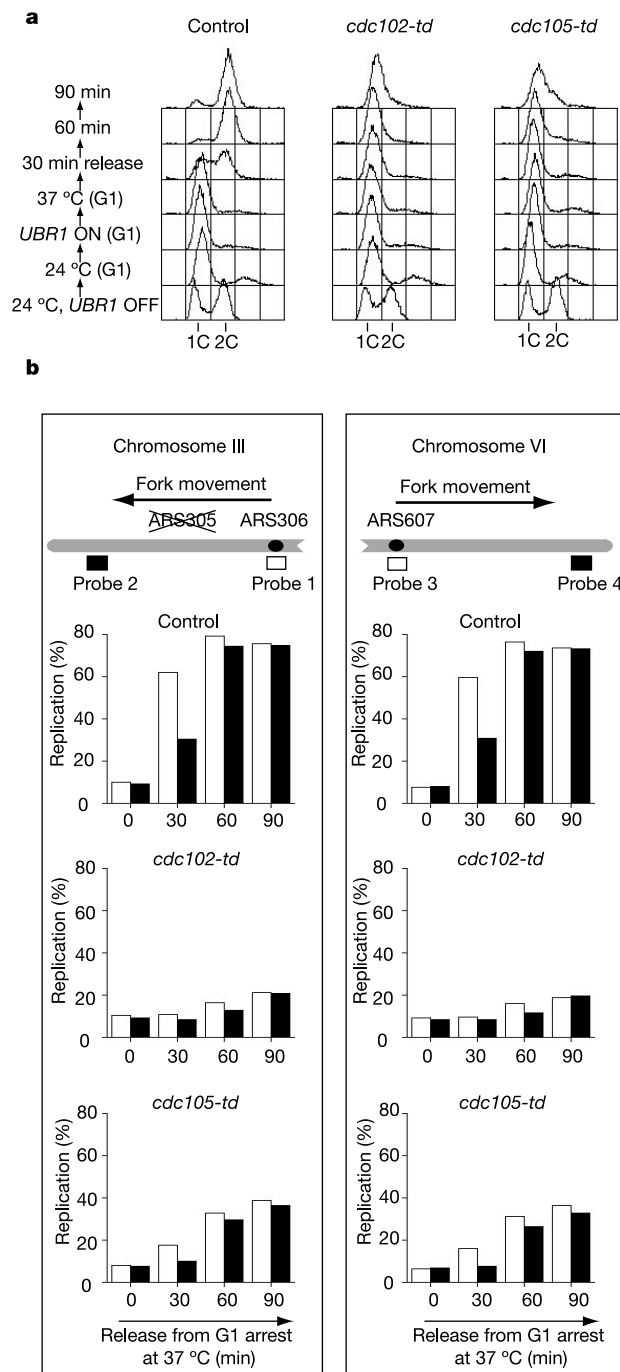
progression of DNA replication forks away from origins, we labelled the same strains as before with heavy isotopes, synchronized cells in G1 phase, and then allowed them to progress into S phase at 24 °C in the presence of hydroxyurea, so that DNA replication forks from early origins stalled within a few kilobases. We then expressed Ubr1 and shifted cells to 37 °C to inactivate Cdc102 or Cdc105, before transferring them into fresh medium lacking hydroxyurea, to determine whether chromosome replication could continue. In the control strain, replication of the genome was essentially complete within 30 min, as detected by flow cytometry (Fig. 4a). In contrast, the continuation of replication was severely impaired after inactivation of Cdc102 or Cdc105 (Fig. 4a; a similar experiment showing even later time points is given in Supplementary Fig. 4), suggesting that the progression of DNA replication forks is defective in their absence. By examining individual replicons we saw, even at 24 °C in hydroxyurea, that the replication of regions containing the

**c**, Yeast extracts from control or *CDC102-TAP* strains were purified sequentially over IgG-Sepharose and calmodulin affinity resin, before detection of Cdc102 and Cdc105 proteins by immunoblotting. **d**, Association of Cdc102–9Myc with the early origins ARS1 and ARS305 was examined in the indicated strains by chromatin immunoprecipitation (IP). HU, hydroxyurea.

early origins ARS306 and ARS607 is relatively defective in the *cdc102-td* and *cdc105-td* degon strains in comparison to the control (Fig. 4b). Replication from early origins under such conditions is, however, sufficient to activate the ‘origin-firing checkpoint’<sup>12</sup>, as mitosis is blocked in all cells and replication of the late origin ARS501 is completely inhibited during hydroxyurea arrest (see Supplementary Fig. 5).

In the control strain, it takes less than 30 min for forks established at ARS306 or ARS607 to reach the end of the replicon (Fig. 4b; compare the replication of probes 1 and 3 or probes 4 and 6). In contrast, although some forks do reach the end of the replicon with apparently similar kinetics in the *cdc102-td* and *cdc105-td* degon strains, a significant proportion is defective (21% of forks from ARS306 and 32% from ARS607 in the *cdc102-td* degon strain; 33% of forks from ARS306 and 37% from ARS607 in the *cdc105-td* degon strain; see Methods). These forks travel slowly through the

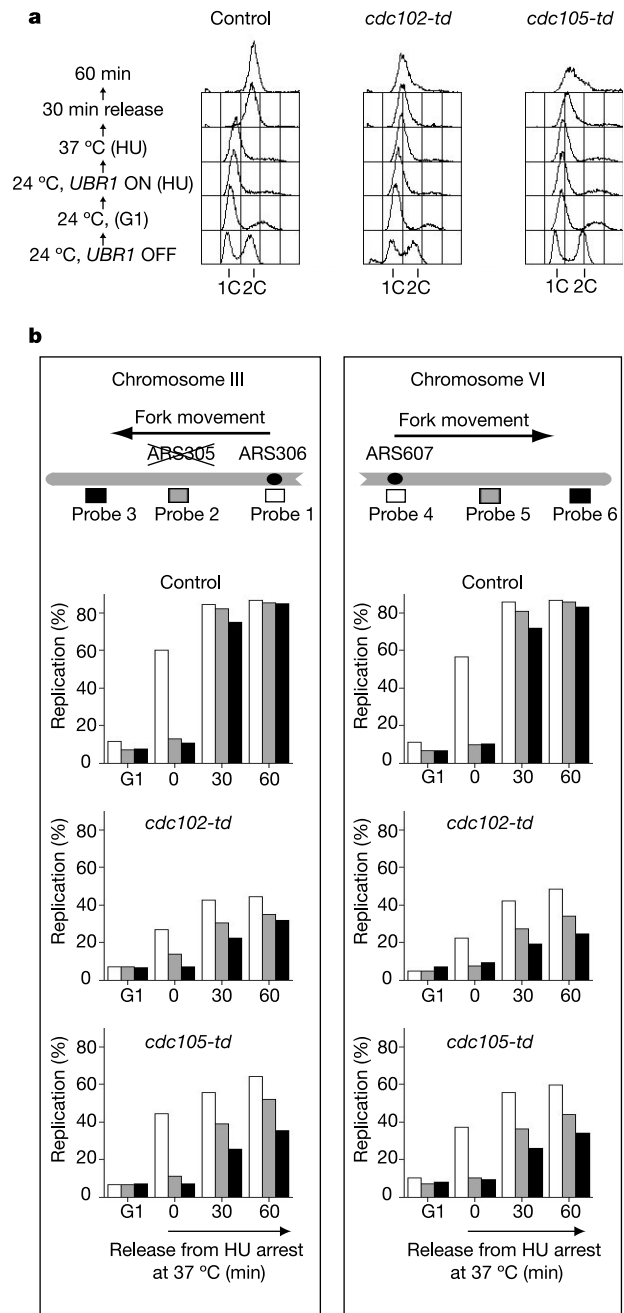




**Figure 3** Cdc102 and Cdc105 are essential for an early step of origin replication. **a**, Flow cytometry was used to measure the DNA content of control, *cdc102-td* and *cdc105-td* strains after release from G1 arrest at 37 °C. **b**, Replication of two individual replicons was examined: at the left end of chromosome III (left panel) and the right end of chromosome VI (right panel).

two replicons, as shown by the progressive decrease in replication with increasing distance from the origin (Fig. 4b; compare probes 2 and 3 and probes 5 and 6). This may reflect a reduced rate of synthesis at the forks or repeated stalling throughout the replicon.

Taken together, these experiments indicate that the complex containing Cdc102 and Cdc105 has an important role in the establishment of DNA replication forks at origins of DNA replication, and is also important for the normal progression of forks away from origins. Future screens of the degron collection will probably allow the identification of new proteins required for many other



**Figure 4** Cdc102 and Cdc105 are important for the normal progression of DNA replication forks away from origins. **a**, DNA content was measured by flow cytometry as before. **b**, Replication of the left end of chromosome III (left panel) and the right end of chromosome VI (right panel) was measured as above.

processes of eukaryotic cell biology. The collection will also facilitate the testing of predictions made by other genomic or proteomic approaches, such as 'global' studies of gene expression, protein localization, protein interactions, and so on<sup>8,9,18–20</sup>.

We note that the degron approach described here for *S. cerevisiae* could, in principle, be applied to other eukaryotes. The degron is unstable in both fission yeast<sup>21</sup> and mouse cells<sup>22</sup> at 37 °C, although efficient degradation through increased expression of Ubr1 has yet to be achieved in these systems. Although genetic modification is clearly much more challenging in higher eukaryotes, it may be possible to use the techniques of RNA interference to replace an endogenous protein with a degron version, so that the immediate consequences of rapidly depleting a protein can be examined. □

## Methods

### Construction of degron strains

Degron strains were made by a one-step PCR approach using a yeast strain in which the only copy of the *UBR1* gene was regulated by the galactose-inducible *GAL1,10* promoter<sup>3</sup> and the degron cassette described in Supplementary Fig. 6. For each essential gene of unknown function, a PCR product was generated using 70-base oligonucleotides beginning with 50 nucleotide homology to either the promoter region or the start of the open reading frame, and ending with 20 nucleotides equivalent to sequences within the degron cassette (see Supplementary Fig. 6; details will be provided on request).

We transformed the PCR products into a haploid *GAL-UBR1* strain or a diploid strain heterozygous for *GAL-UBR1*, and cells were grown at 24 °C on YPD medium containing 0.1 mM CuSO<sub>4</sub>. Integrations were checked by a series of PCR reactions using combinations of oligonucleotides corresponding to sequences either side of the integration site or within the degron cassette. For each essential protein of unknown function we generated two independent haploid clones, and confirmed that they behaved in the same way.

### Growth of degron strains

Asynchronous cultures were grown at 24 °C in YP medium containing 2% raffinose as the carbon source (YP<sub>Raff</sub>) and 0.1 mM CuSO<sub>4</sub>. To inactivate a degron fusion protein, expression of Ubr1 was induced by transferring cells to YP medium containing 2% galactose (YP<sub>Gal</sub>) and 0.1 mM CuSO<sub>4</sub> for 35–60 min at 24 °C, before changing to YP<sub>Gal</sub> medium pre-warmed to 37 °C, and continuing incubation for at least 45–60 min.

### Dense-isotope substitution experiments

These experiments were performed essentially as described previously<sup>3,23</sup>, using strains in which the early origin of DNA replication ARS305 had been deleted. See Supplementary Information for full details of the experimental protocol together with details of the DNA sequences examined.

The defect in fork progression after inactivation of Cdc102 or Cdc105 was calculated using the data in Fig. 4. In control cells, forks from ARS306 or ARS607 take 20–30 min to reach the end of the corresponding replicons on release from hydroxyurea at 37 °C (Fig. 4b and M.K., unpublished data). Therefore, we can estimate the defect in fork progression over any 30-min period by dividing the percentage replication of a fragment close to the end of the replicon (probe 3 for chromosome III; probe 6 for chromosome VI) by the percentage replication of the corresponding origin 30-min earlier (probe 1 for chromosome III; probe 4 for chromosome VI). This was done for the two periods 0–30 min and 30–60 min, and the mean value was determined.

### Construction and use of the *tetO2-MCM4* strain

The chromosomal locus of the *MCM4* gene was modified to replace the endogenous promoter with the doxycycline-regulatable *tetO2* promoter using a one-step PCR method as described previously<sup>5</sup>. For the experiment shown in Fig. 1b, the *mcm4-td* and control strains were grown in YP<sub>Raff</sub> plus 0.1 mM CuSO<sub>4</sub> at 24 °C, before transferring to YP<sub>Gal</sub> at 24 °C for 60 min, and then shifting cells to 37 °C in YP<sub>Gal</sub> medium. The *tetO2-MCM4* strain and its control (both containing Tet-SSN6<sup>5</sup>) were treated identically to the degron strain except that 20 µg ml<sup>-1</sup> doxycycline was added after transferring cells to the medium containing galactose.

### RNA and protein analysis

RNA was prepared from cells using the Qiagen RNeasy mini-prep kit. *MCM4* mRNA was detected using a probe corresponding to nucleotides 1–450 of the *MCM4* open reading frame. Mcm4 protein was detected in immunoblots using an antibody raised against a peptide from the C terminus of the protein (Santa Cruz sc-6685). Cdc102 and Cdc105 were detected using polyclonal antibodies raised against six-histidine-tagged recombinant proteins purified from *Escherichia coli* extracts. A complex containing Cdc102 and Cdc105 was purified from yeast extracts using the tandem affinity purification (TAP) method<sup>24,25</sup>.

### Flow cytometry

We measured DNA content as described previously<sup>26</sup>.

### Chromatin immunoprecipitation

ChIP experiments were performed as described previously using PCR primers specific for ARS305 and flanking sequences<sup>27</sup>. To examine the association of Cdc102–9Myc with ARS1 and adjacent sequences, we used the following PCR primers: chromosome IV (485.5 kb) 5′-AAGCGCCCTGATTGACAAG-3′ and 5′-GCTGGAGGAATTCGAGAATG-3′; chromosome IV (462.5 kb) 5′-TGGTGTGTGATGTAAGCGGAG-3′ and 5′-AAAGTCAACCCCTGCGATG-3′; chromosome IV (466.5 kb) 5′-AGTGCCCTAGAAAGGTGCTAC-3′ and 5′-GGAACACACCGGCAAACT-3′. To show that the Cdc7 kinase is required for association of Cdc102–9Myc with chromatin, *CDC102-9Myc* and *cdc7-1 CDC102-9Myc* strains were grown in YP<sub>Raff</sub> medium at 24 °C, arrested in G1 with α-factor, then incubated at 34 °C with α-factor for 50 min before releasing from G1 arrest at 34 °C in the presence of 0.2 M hydroxyurea. To show that association of Cdc102–9Myc with chromatin is CDK-dependent, *CDC102-9Myc* and *GAL-SIC1ΔNT CDC102-9Myc* strains were grown in YP<sub>Raff</sub> medium at 23 °C, arrested in G1 with α-factor, then transferred for 45 min to YP<sub>Gal</sub> medium containing α-factor to induce expression of Sic1ΔNT<sup>28,29</sup>, before release from G1 arrest at 23 °C in the presence of 0.2 M hydroxyurea.

Received 25 February; accepted 6 May 2003; doi:10.1038/nature01692.

Published online 25 May 2003.

- Kamath, R. S. *et al.* Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* **421**, 231–237 (2003).
- Giaever, G. *et al.* Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418**, 387–391 (2002).
- Labib, K., Tercero, J. A. & Diffley, J. F. X. Uninterrupted MCM2–7 function required for DNA replication fork progression. *Science* **288**, 1643–1647 (2000).
- Dohmen, R. J., Wu, P. & Varshavsky, A. Heat-inducible degron: a method for constructing temperature-sensitive mutants. *Science* **263**, 1273–1276 (1994).
- Belli, G., Gari, E., Aldea, M. & Herrero, E. Functional analysis of yeast essential genes using a promoter-substitution cassette and the tetracycline-regulatable dual expression system. *Yeast* **14**, 1127–1138 (1998).
- Dragon, F. *et al.* A large nucleolar U3 ribonucleoprotein required for 18S ribosomal RNA biogenesis. *Nature* **417**, 967–970 (2002).
- Wu, L. F. *et al.* Large-scale prediction of *Saccharomyces cerevisiae* gene function using overlapping transcriptional clusters. *Nature Genet.* **31**, 255–265 (2002).
- Uetz, P. *et al.* A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*. *Nature* **403**, 623–627 (2000).
- Ito, T. *et al.* A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl Acad. Sci. USA* **98**, 4569–4574 (2001).
- Takayama, Y. *et al.* GINS, a novel multiprotein complex required for chromosomal DNA replication in budding yeast. *Genes Dev.* **17**, 1153–1165 (2003).
- Kubota, Y. *et al.* A novel ring-like complex of *Xenopus* proteins essential for the initiation of DNA replication. *Genes Dev.* **17**, 1141–1152 (2003).
- Santocanale, C. & Diffley, J. F. X. A Mec1- and Rad53-dependent checkpoint controls late-firing origins of DNA replication. *Nature* **395**, 615–618 (1998).
- Yabuki, N., Terashima, H. & Kitada, K. Mapping of early firing origins on a replication profile of budding yeast. *Genes Cells* **7**, 781–789 (2002).
- Bousset, K. & Diffley, J. F. X. The Cdc7 protein kinase is required for origin firing during S phase. *Genes Dev.* **12**, 480–490 (1998).
- Donaldson, A. D., Fangman, W. L. & Brewer, B. J. Cdc7 is required throughout the yeast S phase to activate replication origins. *Genes Dev.* **12**, 491–501 (1998).
- Donaldson, A. D. *et al.* CLB5-dependent activation of late replication origins in *S. cerevisiae*. *Mol. Cell* **2**, 173–183 (1998).
- McCarroll, R. M. & Fangman, W. L. Time of replication of yeast centromeres and telomeres. *Cell* **54**, 505–513 (1988).
- Gavin, A. C. *et al.* Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415**, 141–147 (2002).
- Ho, Y. *et al.* Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* **415**, 180–183 (2002).
- Kumar, A. *et al.* Subcellular localization of the yeast proteome. *Genes Dev.* **16**, 707–719 (2002).
- Lindner, K., Gregan, J., Montgomery, S. & Kearsey, S. E. Essential role of MCM proteins in premeiotic DNA replication. *Mol. Biol. Cell* **13**, 435–444 (2002).
- Levy, F., Johnston, J. A. & Varshavsky, A. Analysis of a conditional degradation signal in yeast and mammalian cells. *Eur. J. Biochem.* **259**, 244–252 (1999).
- Tercero, J. A., Labib, K. & Diffley, J. F. X. DNA synthesis at individual replication forks requires the essential initiation factor, Cdc45p. *EMBO J.* **19**, 2082–2093 (2000).
- Rigaut, G. *et al.* A generic protein purification method for protein complex characterization and proteome exploration. *Nature Biotechnol.* **17**, 1030–1032 (1999).
- Tasto, J. J., Carnahan, R. H., McDonald, W. H. & Gould, K. L. Vectors and gene targeting modules for tandem affinity purification in *Schizosaccharomyces pombe*. *Yeast* **18**, 657–662 (2001).
- Labib, K., Diffley, J. F. X. & Kearsey, S. E. G1-phase and B-type cyclins exclude the DNA-replication factor Mcm4 from the nucleus. *Nature Cell Biol.* **1**, 415–422 (1999).
- Kamimura, Y., Tak, Y.-S., Sugino, A. & Araki, H. Sld3, which interacts with Cdc45 (Sld4), functions for chromosomal DNA replication in *Saccharomyces cerevisiae*. *EMBO J.* **20**, 2097–2107 (2001).
- Drury, L. S., Perkins, G. & Diffley, J. F. X. The Cdc4/34/53 pathway targets Cdc6p for proteolysis in budding yeast. *EMBO J.* **16**, 5966–5976 (1997).
- Desdouts, C. *et al.* Evidence for a Cdc6p-independent mitotic resetting event involving DNA polymerase α. *EMBO J.* **17**, 4139–4146 (1998).

**Supplementary Information** accompanies the paper on [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank J. Diffley for support and encouragement in the early stages of this project and for comments on the manuscript, together with N. Jones. We are grateful to members of our laboratory for helpful discussions, S. Pepper for helping to analyse mass spectrometry data, and to M. Segurado and E. Rawson for their assistance. We thank E. Schiebel and K. Gould for plasmids; E. Schiebel for help in growing fermentor cultures; and H. Araki and Y. Kamimura for assistance with ChIP and for communicating unpublished data, together with H. Takisawa. A.S.-D. thanks F. Perez-Campo for her support. This work was funded by Cancer Research UK, from whom K.L. receives a Senior Cancer Research Fellowship, and by the European Community, from whom A.S.-D. receives a Marie Curie Fellowship.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to K.L. (klabib@picr.man.ac.uk).

# The magnetic field of an isolated neutron star from X-ray cyclotron absorption lines

G. F. Bignami<sup>\*†‡</sup>, P. A. Caraveo<sup>‡</sup>, A. De Luca<sup>‡§</sup> & S. Mereghetti<sup>‡</sup>

<sup>\*</sup> Centre d'Etude Spatiale des Rayonnements, CNRS-UPS, 9 Avenue du Colonel Roche, 31028 Toulouse Cedex 4, France

<sup>†</sup> Università degli Studi di Pavia, Dipartimento Fisica Nucleare e Teorica, Via Bassi 6, 27100 Pavia, Italy

<sup>‡</sup> Istituto di Astrofisica Spaziale e Fisica Cosmica, Sezione di Milano "G. Occhialini", Via Bassini 15, 20133 Milano, Italy

<sup>§</sup> Università di Milano Bicocca, Dipartimento Fisica, Piazza della Scienza 3, 20126 Milano, Italy

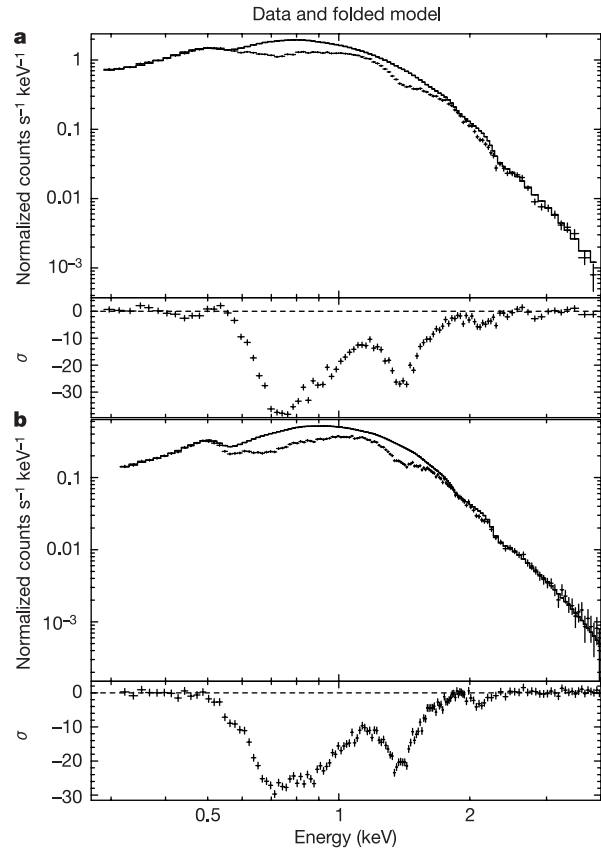
Isolated neutron stars are highly magnetized, fast-rotating objects that form as an end point of stellar evolution. They are directly observable in X-ray emission, because of their high surface temperatures. Features in their X-ray spectra could in principle reveal the presence of atmospheres<sup>1</sup>, or be used to estimate the strength of their magnetic fields through the cyclotron process<sup>2</sup>, as is done for X-ray binaries<sup>3,4</sup>. Almost all isolated neutron star spectra observed so far appear as featureless thermal continua<sup>5,6</sup>. The only exception is 1E1207.4–5209 (refs 7–9), where two deep absorption features have been detected<sup>10,11</sup>, but with insufficient definition to permit unambiguous interpretation. Here we report a long X-ray observation of the same object in which the star's spectrum shows three distinct features, regularly spaced at 0.7, 1.4 and 2.1 keV, plus a fourth feature of lower significance, at 2.8 keV. These features vary in phase with the star's rotation. The logical interpretation is that they are features from resonant cyclotron absorption, which allows us to calculate a magnetic field strength of  $8 \times 10^{10}$  G, assuming the absorption arises from electrons.

In August 2002, the XMM-Newton spacecraft devoted two of its orbits to 1E1207.4–5209 for a total observing time of 257,303 s, the longest observation with the EPIC instrument<sup>12,13</sup> on a galactic source. Selecting photons with energies between 0.2–4 keV, inside a 45" radius from the source, 208,000 photons were found in the PN CCD<sup>12</sup> and 74,600 and 76,700, respectively, in the metal oxide semiconductor CCDs MOS1 and 2 (ref. 13).

Figure 1 shows the spectra of the source as measured by the PN (Fig. 1a) and the two MOS detectors summed together (Fig. 1b). Both independent spectral distributions show, in addition to the two wide absorption features centred at 0.7 and 1.4 keV (refs 10, 11), a third feature around 2.1 keV. The third feature is seen in the three independent PN and MOS1–MOS2 data sets and was also marginally visible in the shorter XMM-Newton observation of December 2001 (ref. 11). The statistical significance of the third feature can now be gauged by the result of an *F*-test. Its inclusion in the overall spectral fit (see Fig. 1 legend) as a gaussian line in absorption yields an improvement in the quality of the fit which has a chance occurrence probability of about  $2 \times 10^{-5}$  for the PN and about  $10^{-3}$  for each of the MOS. A fourth dip at about 2.8 keV is also seen in the PN data and its chance occurrence probability is estimated to be about  $5 \times 10^{-2}$ .

We are aware of the existence of edges of instrumental origin due to Au M (2.209 keV) and Si K (1.839 keV) in the general region of our feature at 2.1 keV. In the spectra of very bright sources, such edges can produce small calibration leftovers (see ref. 14 for example), seen at most at a level of 5%, fully compatible with the EPIC calibration accuracy<sup>15</sup>, with an equivalent width of <10 eV (S. Molendi, personal communication). The ~2.1-keV feature present in our spectra is much stronger (with an equivalent width about 15

times higher and deviations from the continuum model at the level of ~25% as opposed to <5%). Moreover, the full-width at half-maximum (FWHM) of our feature is 370 eV, as opposed to the value of 90 eV measured for the calibration leftovers. These characteristics give us confidence about the reality and the non-instrumental nature of our third feature.



**Figure 1** Spectra collected by the PN (a) and MOS (b) cameras on the EPIC instrument during the observation of 4–6 August 2002. The two MOS cameras were operated in 'full frame' mode<sup>13</sup>, and the PN camera was in 'small window' mode for accurate timing of source photons (6-ms resolution<sup>12</sup>). All cameras used the thin filter<sup>13</sup>. The data were processed with the XMM-Newton Science Analysis Software, version 5.3.3. After removing time intervals with high particle background and correcting for dead time, we obtain a net exposure of 138.4 ks for the PN camera and 193.6 and 194.7 ks for the MOS1 and MOS2 cameras, respectively. Our count rates are slightly larger than those reported<sup>11</sup>, owing to our use of the thin filter:  $1.486 \pm 0.003$  counts s<sup>-1</sup> for the PN,  $0.379 \pm 0.001$  counts s<sup>-1</sup> (MOS1) and  $0.388 \pm 0.001$  counts s<sup>-1</sup> (MOS2). Data points and best-fitting continuum spectral models are shown, together with residuals in units of standard deviations  $\sigma$  from the best-fitting continuum. The best spectral model ( $\chi^2 = 1.25$ , 111 degrees of freedom, d.o.f.) describes the continuum as the sum of two blackbody curves with  $kT = 0.211 \pm 0.001$  keV, for an emitting radius  $R = 2.95 \pm 0.05$  km (assuming a distance  $d = 2.2$  kpc; ref. 26), and  $0.40 \pm 0.02$  keV ( $R = 250 \pm 50$  m). The line of sight column density,  $N_H$  is  $(1.0 \pm 0.1) \times 10^{21}$  cm<sup>-2</sup>, to be compared with  $1.6 \times 10^{21}$  cm<sup>-2</sup> estimated from radio observations of the associated supernova remnant G296.5 + 10.0 (ref. 26). Any other model for the continuum fails to give such a good fit to the data: we tried a composite (blackbody + power-law) model ( $\chi^2 = 1.9$ , 111 d.o.f.) and a hydrogen atmosphere model<sup>19</sup> ( $\chi^2 = 1.7$ , 113 d.o.f.). The central energies of the three absorption features are  $0.72 \pm 0.02$  keV,  $1.37 \pm 0.02$  keV and  $2.11 \pm 0.03$  keV. The possible fourth feature in the PN spectrum is at  $2.85 \pm 0.06$  keV. The optical depths  $\tau$  for the two main features are similar ( $\tau_1 = 0.54 \pm 0.02$ ;  $\tau_2 = 0.52 \pm 0.02$ ); however, the integrated number of scattered photons is more than four times higher for the feature at 0.7 keV. The third feature is less intense. All errors in the spectral parameters are at the 90% confidence level.



The three absorption features shown in Fig. 1 are wider than the CCD energy resolution and their intensities are definitely different. No significant substructure was detected in any of them. The shapes of the two main features cannot be described by any single gaussian profile. Detailed analytical or numerical fits to the feature shapes are beyond the scope of this work, but we note that at least two gaussian profiles, always larger than the instrument resolution, are needed to describe each of the features at 0.7 and 1.4 keV (see also Fig. 1 legend). In any case, Doppler broadening by thermal electron motion in the  $kT \approx 0.2$  keV environment is smaller than the CCD energy resolution. Unfortunately, the higher spectral resolution of the reflection grating spectrometer (RGS)<sup>16</sup> cannot provide any additional clue on this source because its limited sensitivity requires a drastic rebinning of the dispersed data. The resulting spectrum is coarser than those shown in Fig. 1.

After the first EPIC observation, variations of the lines as a function of the pulsar phase were reported<sup>11</sup>. To investigate such an effect in our longer data set, we first had to search for the best source period. This was done following the steps outlined in ref. 11 and yielded a pulsar period  $P = 0.42413076 \pm 0.00000002$  s. The light curve for the total energy range (0.2–4 keV), shown in Fig. 2, is characterized by a nearly sinusoidal modulation with a pulsed fraction of about 7%. In it, four phase intervals can easily be identified (see Fig. 2).

Figure 3 shows the four colour-coded spectra corresponding to the four phase intervals. The bulk of the phase dependence of the spectra is seen to be due to the absorption lines variation, while the continua, as seen in the spectral regions not affected by the lines, remain reasonably constant. Indeed, we could go as far as saying that the X-ray source pulsation is largely due to the phase variation of the lines with the neutron star rotation. To our knowledge, this is the first time that such a phenomenon has been observed for a rotating compact object. The inset to Fig. 3 allows for a direct appreciation of the absorption feature phase variation, as seen through the spectral residuals.

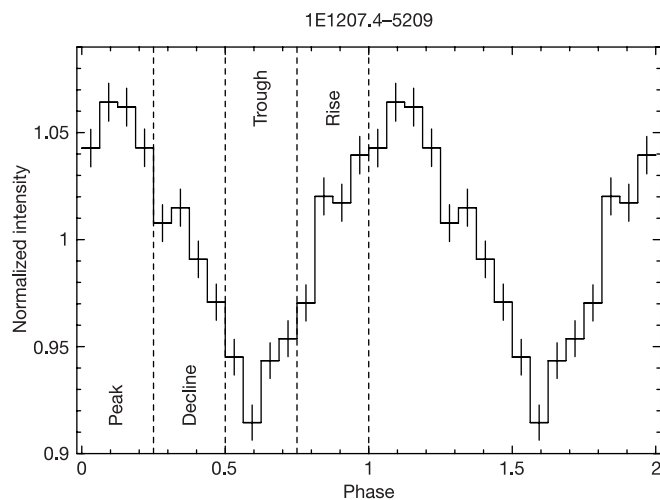
To summarize, the soft X-ray spectrum of 1E1207.4–5209 is characterized by three absorption features with integrally spaced energies (0.7, 1.4 and 2.1 keV; there may even be a fourth one at  $\sim 2.8$  keV), and which are varying in phase with the neutron star rotation, possibly accounting for the bulk of

the X-ray pulsation. They also have, in our deep XMM-Newton observation, an extent, or depth, which appears to be globally decreasing with energy (see Fig. 1), although this effect appears to be phase-dependent.

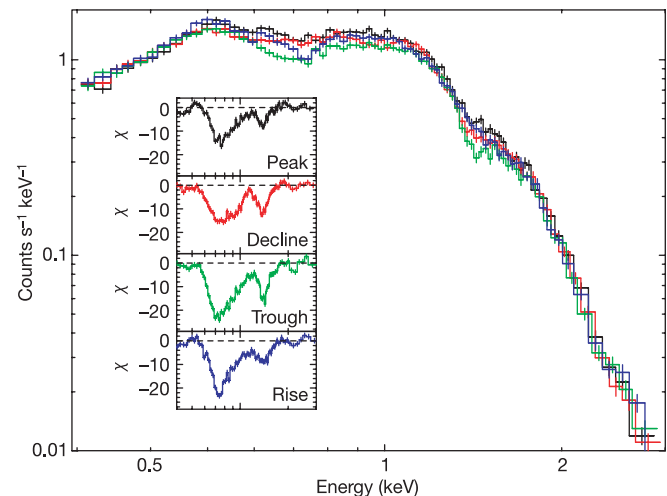
X-ray spectral features from an isolated neutron star (INS) can be due to atomic transition lines, photoabsorption edges or cyclotron lines. The atmosphere of an INS is a very difficult environment for physics, owing to the necessity of working with an *a priori* unknown nuclear species composition and level of ionization, in a Landau regime of high magnetic field. We refer to extensive work<sup>17,18</sup> on possible transition lines or photoabsorption edges. Their final preference for explaining the Chandra features in 1E1207.4–5209 with He-like mid-Z (O or Ne) absorption lines, although certainly self-consistent, seems no longer justified by the present XMM-Newton data.

Atmospheric absorption models<sup>1,19</sup> also meet with the strong difficulty of the lack of similar features in any of the several INSs for which good Chandra/XMM-Newton spectra are now available. Data for such diverse INSs as Vela, pulsars PSR B0656+14 and PSR B1055–52, the millisecond pulsar PSR J0437–4715 and the two ‘dim’ INSs RX J1856.5–3754 and RX J0720.4–3125 (see recent reviews<sup>5,6</sup>) were explored and none shows absorption or emission features. These objects cover a vast range in age (from  $10^4$  yr to ‘eternity’ in the case of the millisecond pulsar), in magnetic fields  $B$ , in possible interstellar accretion rate, and so on. It is difficult to imagine none of these INSs showing a phenomenology similar to 1E1207.4–5209, if the features of the latter were due to atmospheric absorption.

Occam’s razor, in the end, leads us to conclude that cyclotron resonance scattering explains the features presented above: three lines, correctly spaced by integral factors (within the uncertainties on their measured centroids) and with a phase variation naturally following the pulsar  $B$ -field rotation, as observed for the much brighter accreting neutron stars in binary systems, see refs 20, 21 for example.



**Figure 2** Total light curve relative to 208,000 photons collected by the PN camera folded at the best period of  $P = 0.42413076$  s. When compared to the Chandra measurement of January 2000 (refs 9, 30), this yields a refined estimate of the pulsar spindown of  $\dot{P} = (1.4 \pm 0.3) \times 10^{-14}$  s s<sup>-1</sup>. The four equal phase intervals used to assess the spectral variation (namely, ‘peak’, ‘decline’, ‘trough’ and ‘rise’) are also shown.



**Figure 3** Comparison of the four PN spectra for phase intervals as defined in Fig. 2. Note that absolute (counts s<sup>-1</sup> keV<sup>-1</sup>) spectra are plotted. Colours indicate spectral variation: black, peak; red, decline; green, trough; blue, rise. The peak of the total light curve corresponds to the phase interval where the absorption lines are at their minimum (black data points) while the light-curve trough happens when the absorption lines are more important (green data points). Inset, the four panels show the residuals of the phase-dependent spectra from the two-blackbody fit used in Fig. 1a. Phase variations in both shape and intensity for all features are apparent. The comparison of phase-dependent residuals with Fig. 3 of ref. 11 shows the improvement achieved with the present long observation.

Note that small deviations from the 1:2:3 ratio of the line centroids can also be induced by the complex geometrical and physical environment related to harmonic generation<sup>22</sup>. Our best-fitting temperature for the source continuum emission ( $\sim 0.21$  keV) is nicely consistent with the equilibrium Compton temperature due to resonant scattering, computed<sup>2</sup> to be 0.27 of the energy of the first harmonic, or about 0.19 keV.

Following ref. 23, we note that in cyclotron scattering above the surface of an INS, the cross-section (and the relative strength of the harmonic lines) depends on the angle between the observer and the magnetic field. This provides a natural explanation for the phase variation of the feature depths. There are also compelling physical reasons<sup>23,24</sup> why all line widths should vary in phase, with maximum broadening appearing when the observer's line of sight and the pulsar  $B$  field are aligned. Of course, such asymmetric and varying line profiles render problematic any precise evaluation of each line's central energy as well as their harmonic spacings. Thus, *a fortiori*, considerations on line shapes and their centroid energies based on the phase-averaged spectra of Fig. 1 have little physical meaning. Rather, any physical interpretative work on 1E1207.4–5209 should be driven by phase-resolved spectroscopy.

If the  $\sim 0.7$ -keV line is the fundamental electron cyclotron frequency, then  $B$  averaged over the star's disk is around  $6 \times 10^{10}$  (1 +  $z$ ), or  $8 \times 10^{10}$  G, assuming a 'standard' 20% gravitational redshift,  $z$ . For protons, the field should be higher by a factor equal to the proton-to-electron mass ratio, yielding  $B = 1.6 \times 10^{14}$  G, that is, a magnetar-type magnetic field<sup>25</sup>. Neither of these values agrees with the  $B$  field inferred from the 1E1207.4–5209 timing parameters, which give  $(2\text{--}3) \times 10^{12}$  G under the rotating dipole hypothesis. This hypothesis, however, is not problem-free for our object because the current period derivative,  $\dot{P} = (1.4 \pm 0.3) \times 10^{-14} \text{ s s}^{-1}$  implies a pulsar age of about  $4.8 \times 10^5$  yr, which is totally incompatible with that of the supernova remnant associated with it, no more than about  $10^4$  yr old<sup>26</sup>. Although the reported tentative radio detection<sup>27</sup> may help to monitor the pulsar timing parameters better, it seems unlikely that improved timing could solve the age inconsistency problem. More drastic assumptions, such as low spin frequency at birth, should possibly be considered.

To reconcile the pulsar period derivative and the  $B$  value inferred in the case of electron cyclotron emission close to the neutron star surface, one should invoke additional pulsar braking mechanisms, such as the debris disk configuration proposed<sup>28</sup>. Of course, electron cyclotron scattering at a distance  $R \approx 3\text{--}4$  stellar radii (that is, 20–30 km above the neutron star surface) would fit all the observations.

Alternatively, the cyclotron features could be due to protons. This would require an improbable surface magnetic field about 100 times stronger than that deduced from the spindown. Although a single proton cyclotron feature could be present in the optical spectrum of Geminga<sup>29</sup>, the  $B$ -field requirements for 1E1207.4–5209 seem too large to support the proton cyclotron explanation for the observed X-ray features.

On the whole, electron resonance scattering in an abnormally low magnetic field seems to explain the current XMM-Newton observations best. The cyclotron nature of the features, as opposed to the atmospheric one, would also naturally explain the uniqueness of the 1E1207.4–5209 spectrum in the INS context, because the powerful combination of throughput and energy resolution of XMM-Newton/EPIC is only effective for less than a decade in photon energy. Most INSs, with a more normal  $B$  field, may have electron cyclotron features in the 10–20 keV range, that is, where such features were observed for the much brighter accreting objects. □

Received 3 February; accepted 6 May 2003; doi:10.1038/nature01703.

- Zavlin, V. E. & Pavlov, G. G. *Modeling Neutron Star Atmospheres* (eds Becker, W., Lesch, H. & Trümper, J.) 263–272 (MPE Report 278, Max Plank Institut für Extraterrestrische Physik, Garching bei, München, 2002).
- Lamb, D. Q., Wang, J. C. L. & Wasserman, I. M. Energetics and dynamics of resonant and nonresonant scattering in strong magnetic fields. *Astrophys. J.* **363**, 670–693 (1990).
- Trümper, J. *et al.* Evidence for strong cyclotron line emission in the hard X-ray spectrum of Hercules X-1. *Astrophys. J.* **219**, L105–L110 (1978).
- Weather, Wm. A. *et al.* An absorption feature in the spectrum of the pulsed hard X-ray flux from 4U0115+63. *Nature* **282**, 240–243 (1979).
- Pavlov, G. G., Zavlin, V. E. & Sanwal, D. *Thermal Radiation from Neutron Stars: Chandra Results* (eds Becker, W., Lesch, H. & Trümper, J.) 273–289 (MPE Report 278, Max Plank Institut für Extraterrestrische Physik, Garching bei, München, 2002).
- Becker, W. & Aschenbach, B. in *X-ray Observations of Neutron Stars and Pulsars: First Results from XMM-Newton* (eds Becker, W., Lesch, H. & Trümper, J.) 64–86 (MPE Report 278 Max Plank Institut für Extraterrestrische Physik, Garching bei, München, 2002).
- Bignami, G. F., Caraveo, P. A. & Mereghetti, S. On the optical counterpart of 1E1207.4–5209, the central X-ray source of a ring-shaped supernova remnant. *Astrophys. J.* **389**, L67–L69 (1992).
- Caraveo, P. A., Bignami, G. F. & Trümper, J. Radio-silent isolated neutron stars as a new astronomical reality. *Astron. Astrophys. Rev.* **7**, 209–216 (1996).
- Zavlin, V. E., Pavlov, G. G., Sanwal, D. & Trümper, J. Discovery of 424 millisecond pulsations from the radio-quiet neutron star in the supernova remnant PKS1209–51/52. *Astrophys. J.* **540**, L25–L28 (2000).
- Sanwal, D., Pavlov, G. G., Zavlin, V. E. & Teter, M. A. Discovery of absorption features in the X-ray spectrum of an isolated neutron star. *Astrophys. J.* **574**, L61–L64 (2002).
- Mereghetti, S. *et al.* Pulse phase variations of the X-ray spectral features in the radio-quiet neutron star 1E1207–5209. *Astrophys. J.* **581**, 1280–1285 (2002).
- Strüder, L. *et al.* The European photon imaging camera on XMM-Newton: The pn-CCD camera. *Astron. Astrophys.* **365**, L18–L26 (2001).
- Turner, M. *et al.* The European photon imaging camera on XMM-Newton: The MOS cameras. *Astron. Astrophys.* **365**, L27–L35 (2001).
- Miller, J. M. *et al.* XMM-Newton spectroscopy of the accretion-driven millisecond X-ray pulsar XTE J1751–305 in outburst. *Astrophys. J.* **583**, L99–L102 (2003).
- Kirsch, M. (on behalf of the EPIC collaboration) *EPIC Calibration Issues in Progress Document XMM-SOC-CAL-TN-0018* (EPIC Status of Calibration and Data Analysis), XMM-Newton Science Operation Centre, Villafranca del Castillo, 2002; (<http://xmm.vilspa.esa.es/docs/documents/CAL-TN-0018-2-0.pdf>).
- Den Herder, J. W. *et al.* The reflection grating spectrometer on board XMM-Newton. *Astron. Astrophys.* **365**, L7–L17 (2001).
- Mori, K. *et al.* X-ray spectroscopy of the isolated neutron star 1E1207.4–5209: Atmospheric composition and equation of state. *Astrophys. J.* (submitted); preprint at (<http://arXiv.org/astro-ph/0301161>) (2003).
- Hailey, C. J. & Mori, K. Evidence of a mid-atomic number atmosphere in the neutron star 1E1207.4–5209. *Astrophys. J.* **578**, L133–L136 (2002).
- Zavlin, V. E., Pavlov, G. G. & Shibano, Yu. Model neutron star atmospheres with low magnetic fields. I. Atmospheres in radiative equilibrium. *Astron. Astrophys.* **315**, 141–152 (1996).
- Voges, W. *et al.* Cyclotron lines in the hard X-ray spectrum of Hercules X-1. *Astrophys. J.* **263**, 803–813 (1982).
- Nagase, F. *et al.* Cyclotron line features in the spectrum of the transient X-ray pulsar X0115+634. *Astrophys. J.* **375**, L49–L52 (1991).
- Isenberg, M., Lamb, D. Q. & Wang, J. C. L. Effects of the geometry of the line-forming region on the properties of cyclotron resonant scattering lines. *Astrophys. J.* **505**, 688–714 (1998).
- Freeman, P. E. *et al.* Resonant cyclotron radiation transfer model fits to spectra from gamma-ray burst GRB 870303. *Astrophys. J.* **524**, 772–792 (1999).
- Meszaros, P. & Nagel, W. X-ray pulsar models. I—Angle-dependent cyclotron line formation and comptonization. *Astrophys. J.* **298**, 147–160 (1985).
- Zane, S. *et al.* Proton cyclotron features in thermal spectra of ultramagnetized neutron stars. *Astrophys. J.* **560**, 384–389 (2001).
- Giacani, E. B. *et al.* The interstellar matter in the direction of the supernova remnant G296.5 + 10.0 and the central X-ray source 1E1207.4–5209. *Astron. J.* **119**, 281–291 (2000).
- Camilo, F. Deep searches for young pulsars. *ASP Conf. Ser.* (eds Bailes, M., Nice, D. J. & Thorsett, S.) (Astronomical Society of the Pacific, San Francisco, in the press); preprint at (<http://arXiv.org/astro-ph/0210620>).
- Xu, R. X., Wang, H. G. & Qiao, G. J. A note on the discovery of absorption features in 1E1207.4–5209. *Chin. Phys. Lett.* **20**, 314 (2003); preprint at (<http://arXiv.org/astro-ph/0207079>).
- Bignami, G. F. *et al.* Multiwavelength data suggest a cyclotron feature on the hot thermal continuum of Geminga. *Astrophys. J.* **456**, L111–L114 (1996).
- Pavlov, G. G., Zavlin, V. E., Sanwal, D. & Trümper, J. 1E 1207.4–5209: The puzzling pulsar at the center of the supernova remnant PKS 1209–51/52. *Astrophys. J.* **569**, L95–L98 (2002).

**Acknowledgements** We thank F. Jansen and all the XMM-Newton team in Vilspa for approving and carrying out this target of opportunity observation at very short notice. We also thank W. Becker and M. Turner for helping us to optimize EPIC settings. G.F.B. acknowledges discussions with D. Lamb, G. Ricker and J. Doty about the cyclotron line interpretation of the spectral features of this object. The XMM-Newton data analysis is supported by the Italian Space Agency (ASI). A.D.L. acknowledges an ASI fellowship.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to G.F.B. ([bignami@cesr.fr](mailto:bignami@cesr.fr)).

# Contemporaneous formation of chondrules and refractory inclusions in the early Solar System

Shoichi Itoh & Hisayoshi Yurimoto

Department of Earth and Planetary Sciences, Tokyo Institute of Technology, Meguro, Tokyo 152-8551, Japan

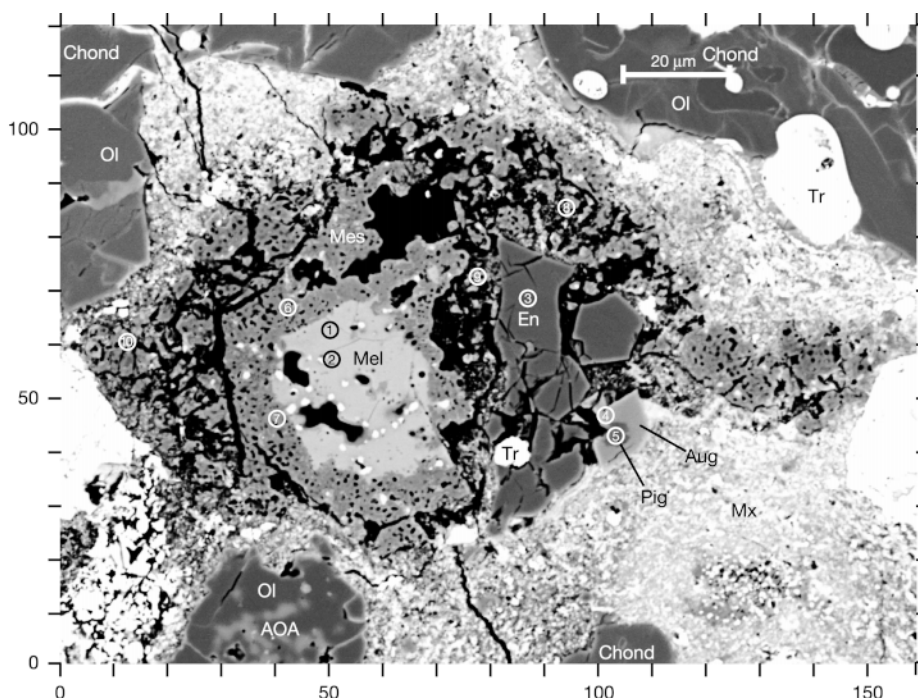
**Chondrules and calcium-aluminium-rich inclusions (CAIs) are preserved materials from the early history of the Solar System, where they resulted from thermal processing of pre-existing solids during various flash heating episodes which lasted for several million years<sup>1</sup>. CAIs are believed to have formed about two million years before the chondrules<sup>2–5</sup>. Here we report the discovery of a chondrule fragment embedded in a CAI. The chondrule's composition is poor in <sup>16</sup>O, while the CAI has a <sup>16</sup>O-poor melilite (Ca, Mg, Al-Silicate) core surrounded by a <sup>16</sup>O-rich igneous mantle. These observations, when combined with the previously reported CAI-bearing chondrules<sup>6–9</sup>, strongly suggest that the formation of chondrules and CAIs overlapped in time and space, and that there were large fluctuations in the oxygen isotopic compositions in the solar nebula probably synchronizing astrophysical pulses.**

The chondrule-bearing CAI, which we have designated A5, was found in the Y-81020 CO3.0 chondrite (Fig. 1). The CAI is about

100  $\mu\text{m}$  across, and consists of a large central polycrystalline melilite (10–13%  $\text{Ak}_{10-13}$ ) clast and a three-pyroxene assemblage enclosed by a porous glassy mesostasis (Supplementary Fig. 1). The mesostasis consists of Al-rich clinopyroxene filaments embedded in Al-rich glass. These textures indicate that the mesostasis was quenched from a liquid. The pyroxene assemblage includes troilite and metal, as well as enstatite, augite and pigeonite. This assemblage has not previously been observed in CAIs<sup>10</sup>, and the chemical compositions of the pyroxene phases (Table 1) are also not known from CAIs. However, such an assemblage is typical of chondrules<sup>11</sup>. Thus, it is evident that the pyroxene assemblage is a fragment of a chondrule, which has mineralogical and textural properties similar to the most frequently observed type-I (that is, low FeO) porphyritic chondrules in the CO chondrites.

The oxygen isotope data are bimodally distributed on the carbonaceous chondrite anhydrous minerals (CCAM) line (Fig. 2). All the <sup>16</sup>O-rich values are from the mesostasis. The O isotope values in both the melilite clast and the pyroxene assemblage are depleted in <sup>16</sup>O. We emphasize that the O isotopic composition of the pyroxenes falls within the range observed in low FeO (type-I) CO chondrules<sup>12,13</sup>. The non-mass-dependent O-isotope distribution indicates that the large melilite crystals in the CAI, and the chondrule fragment (that is, enstatite, pigeonite and augite), did not crystallize from or equilibrate with the surrounding mesostasis liquid.

This non-equilibrium state is also supported by the elemental composition of the mesostasis. The mesostasis composition (Table 1 and Supplementary Table 1) corresponds to the type-C CAI field of Stolper's diagram<sup>10,14</sup>. This composition shows that spinel and then



**Figure 1** Back-scattered electron image of the chondrule-bearing CAI. The object, named A5, was found in a thin section (no. 56-4 from the National Institute of Polar Research, Tokyo) of the Y81020 CO3.0 chondrite. Horizontal and vertical scales are in  $\mu\text{m}$ . Scale bar also at top right. Locations of SIMS analyses are shown by numbers corresponding to those in Table 1. The circle size round each number shows the approximate analytical area of O isotopic composition by SIMS. The CAI consists of a large central polycrystalline melilite ( $\text{Ak}_{10-13}$ , Mel) clast in which individual melilite crystals are about 20  $\mu\text{m}$  across. Fine perovskite grains (white spots) are scattered in and around the melilite. The melilite clast is enclosed by a porous mesostasis (Mes) composed of a fine-grained mix of Al-rich clinopyroxene (lighter area; MgO, 13.3 wt%;  $\text{Al}_2\text{O}_3$ , 18.1;  $\text{SiO}_2$ , 37.5; CaO, 24.2;  $\text{TiO}_2$ ,

6.3; other elements <0.1) and Al-rich glass (darker area; MgO, 7.0–10.6 wt%;  $\text{Al}_2\text{O}_3$ , 41.2–43.9;  $\text{SiO}_2$ , 24.6–33.3; CaO, 11.7–16.7;  $\text{TiO}_2$ , 0.6–1.9;  $\text{Cr}_2\text{O}_3$ , 0.3–1.0; FeO, 0.3–1.2; other elements <0.1). Also enclosed by the porous mesostasis is a pyroxene assemblage dominated by enstatite (En) with minor pigeonite (Pig) and augite (Aug). The augite is an overgrowth on the pigeonite. Troilite (Tr) is minor phase associated with the pyroxene assemblage. Troilite and Fe-Ni metal (white spots) are scattered in the mesostasis near the pyroxene assemblage. The right lower side of the CAI is broken on surface where the pyroxene assemblage directly borders the matrix (Mx). Other abbreviations: Ol, olivine; Chond, chondrule; AOA, amoeboid olivine aggregate.



anorthite would be the first phases to crystallize, but these are not present in A5. The absence of spinel in the mesostasis implies that the mesostasis quenched at  $>1,550^\circ\text{C}$  (ref. 14). It is believed that the type-C CAIs result from melting precursor refractory condensates from the nebular gas<sup>10</sup>, giving rise to the  $^{16}\text{O}$ -rich composition of the mesostasis. Thus the  $^{16}\text{O}$ -rich mesostasis directly shows that the melting process related to CAI formation.

If CAIs formed before chondrules, then the mesostasis in A5 melted again during the chondrule-forming event. If this sequence of events occurred, CAIs and chondrules should have coexisted in the chondrule-forming region and should have melted by the second melting event of the mesostasis. However, igneous CAIs are rare in CO chondrites<sup>15</sup>. Most of CAIs in CO chondrites were formed by condensation of solids<sup>10</sup>.

The quench texture of the mesostasis suggests that the cooling rate of the CAI was faster than those of typical coarse-grained CAIs<sup>16</sup>. Small O isotopic zoning in the rim of the melilite clast (Fig. 2, Supplementary Figs 2 and 3, Supplementary Table 2) indicates that the melting event partly exchanged O isotopic composition of the melilite with the  $^{16}\text{O}$ -rich mesostasis melt by O self-diffusion. The cooling rate can be evaluated by the O isotopic zoning (Supplementary Fig. 3) and the O self-diffusion coefficient of  $\text{Åk}_{12}$  (ref. 17). According to the compensation timescale<sup>18</sup> and assuming a non-moving surface, the cooling rate,  $s$  (in  $\text{K s}^{-1}$ ), is given by:  $\log s = [(-1.8 \times 10^4)/T_0] + 8.8$  along the  $a$  axis of melilite, and  $\log s = [(-1.6 \times 10^4)/T_0] + 6.8$  along the  $c$  axis, where  $T_0$  is absolute temperature at the start of diffusion. Based on these parameters, the cooling rate of the chondrule-bearing CAI was several hundreds to several tens of degrees per hour if  $T_0 = 1,823\text{ K}$  (estimated liquidus temperature of the mesostasis), indicating a faster cooling rate than that of most CAIs<sup>16</sup>.

Because the melilite clast is completely surrounded by  $^{16}\text{O}$ -rich mesostasis, the  $^{16}\text{O}$ -poor values of the melilite are plausibly the

original O-isotopic composition of the melilite clast. The preservation of the original O-isotopic compositions is consistent with the observation that the CO3.0 chondrites are remarkably pristine, having experienced very little thermal metamorphism or aqueous alteration<sup>15,19,20</sup>. In particular, no evidence of O isotopic disturbance has been observed in fine-grained melilite from CO3.0 chondrites<sup>21</sup>. The large grain size of melilite in the clast (Fig. 1) and preservation of  $^{16}\text{O}$ -rich glass in the mesostasis next to the  $^{16}\text{O}$ -poor melilite clast (Fig. 1, Table 1) also support the suggestion that the O isotopic composition of the melilite clast has not been disturbed after mesostasis formation.

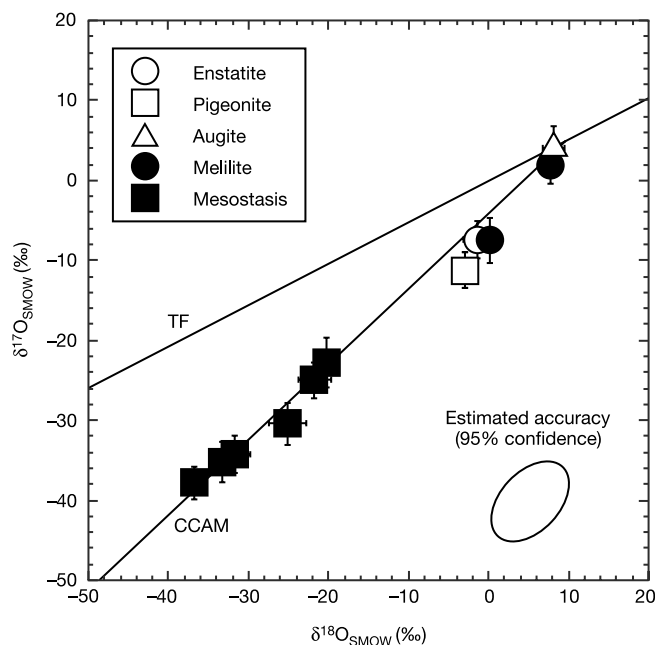
On the basis of these petrological and isotopic results, we conclude that the A5 CAI contains a fragment of a chondrule. This chondrule fragment survived the last melting event in the formation of the CAI, probably because of the short duration of heating and cooling. The chondrule fragment must have formed before the last melting episode of the CAI, which indicates that the formation of the chondrule and the formation of the CAI overlapped, at least partly, both in time and in space. Another important finding in the CAI is that a  $^{16}\text{O}$ -rich liquid encloses a relict CAI melilite fragment with a more  $^{16}\text{O}$ -poor composition. This indicates that O-isotopic variation in CAIs occurred in the solar nebula, and did not monotonically change from  $^{16}\text{O}$ -rich to  $^{16}\text{O}$ -poor. Therefore, the O-isotopic composition of solids in the solar nebula did not uniformly evolve from  $^{16}\text{O}$ -rich to  $^{16}\text{O}$ -poor.

Oxygen is the most abundant element in solid phases and is third in abundance, after H and He, in the solar nebula gas. Thus O isotopes are appropriate tracers of material evolution and global changes in the solar nebula. The O isotopic variation of the chondrule-bearing CAI may be understandable within the framework of the X-wind model<sup>22</sup>: this model offers a mechanism for the bipolar jet and X-ray flare of proto-stars<sup>22</sup>, for contemporaneous formation of CAIs and chondrules around the inner edge of the

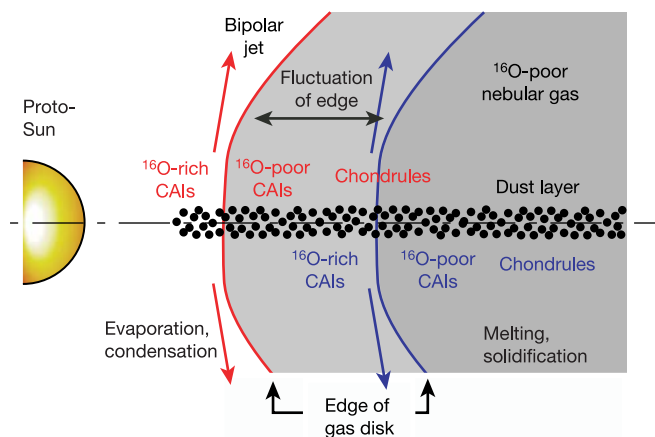
Table 1 Oxygen isotopic compositions (‰) and chemical compositions in A5 inclusion

| Phase                               | Melilite |      |       |       |            | Mesostasis           |       |       |             |       |      |       |
|-------------------------------------|----------|------|-------|-------|------------|----------------------|-------|-------|-------------|-------|------|-------|
| No. in Fig. 1                       | 1        | 2    | 3     | 4     | 5          | 6                    | 6     | 6     | 7           | 8     | 9    | 10    |
| Beam overlap                        | None     | None | None  | None  | Augite 10% | None                 | None  | None  | Melilite 5% | None  | –    | None  |
| Oxygen isotopic compositions (‰)    |          |      |       |       |            |                      |       |       |             |       |      |       |
| $\delta^{17}\text{O}_{\text{SMOW}}$ | –7.5     | 2.0  | –7.4  | 4.1   | –11.2      | –34.3                | –22.8 | –37.8 | –25.0       | –30.4 | –    | –35.2 |
| $\sigma_{\text{mean}}$              | 2.9      | 2.5  | 2.4   | 2.6   | 2.3        | 2.3                  | 3.1   | 2.1   | 2.3         | 2.6   | –    | 2.6   |
| $\delta^{18}\text{O}_{\text{SMOW}}$ | 0.2      | 7.7  | –1.4  | 8.2   | –2.9       | –31.8                | –20.2 | –36.8 | –21.7       | –25.1 | –    | –33.3 |
| $\sigma_{\text{mean}}$              | 1.2      | 1.4  | 1.8   | 1.4   | 1.4        | 2.1                  | 1.2   | 1.3   | 2.1         | 2.4   | –    | 1.6   |
| Chemical compositions (wt %)        |          |      |       |       |            |                      |       |       |             |       |      |       |
| MgO                                 | 2.0      | 1.7  | 38.6  | 21.0  | 34.4       |                      |       | 7.1   | 7.0         | 10.0  | 15.3 | 12.8  |
| $\text{Al}_2\text{O}_3$             | 32.8     | 34.8 | 1.2   | 2.4   | 1.5        |                      |       | 31.9  | 32.4        | 26.7  | 24.5 | 32.0  |
| $\text{SiO}_2$                      | 25.2     | 22.8 | 59.5  | 55.1  | 55.4       |                      |       | 23.0  | 28.0        | 22.9  | 34.8 | 24.4  |
| $\text{K}_2\text{O}$                | ND       | ND   | ND    | ND    | ND         |                      |       | ND    | ND          | ND    | ND   | 0.2   |
| CaO                                 | 42.7     | 40.1 | 0.6   | 20.5  | 3.3        |                      |       | 14.4  | 17.8        | 8.7   | 9.6  | 9.7   |
| $\text{TiO}_2$                      | ND       | ND   | ND    | ND    | 0.1        |                      |       | 0.9   | 1.8         | 1.8   | 0.8  | 1.8   |
| $\text{Cr}_2\text{O}_3$             | ND       | ND   | ND    | 0.6   | 0.3        |                      |       | 0.3   | ND          | 0.7   | 0.4  | ND    |
| $\text{FeO}^*$                      | ND       | ND   | 0.6   | 0.8   | 2.1        |                      |       | 1.0   | 2.2         | 8.7   | 1.4  | 13.4  |
| Ni*                                 | ND       | ND   | ND    | ND    | ND         |                      |       | ND    | ND          | 0.9   | ND   | 0.9   |
| S                                   | ND       | ND   | ND    | ND    | ND         |                      |       | ND    | ND          | 0.4   | ND   | 0.2   |
| Total                               | 102.7    | 99.4 | 100.5 | 100.4 | 97.1       |                      |       | 78.7  | 89.2        | 79.4  | 86.8 | 94.2  |
| Chemical formulae                   |          |      |       |       |            |                      |       |       |             |       |      |       |
| O                                   | 7        | 7    | 3     | 6     | 3          | Stolper's mode (wt%) |       |       |             |       |      |       |
| Mg                                  | 0.13     | 0.11 | 0.96  | 1.11  | 0.90       |                      |       |       |             |       |      |       |
| Al                                  | 1.72     | 1.88 | 0.02  | 0.10  | 0.03       |                      |       |       |             |       |      |       |
| Si                                  | 1.12     | 1.05 | 0.99  | 1.96  | 0.97       |                      |       |       |             |       |      |       |
| Ca                                  | 2.04     | 1.97 | 0.01  | 0.78  | 0.06       |                      |       |       |             |       |      |       |
| Ti                                  | ND       | ND   | ND    | ND    | 0.00       |                      |       |       |             |       |      |       |
| Cr                                  | ND       | ND   | ND    | 0.02  | 0.00       |                      |       |       |             |       |      |       |
| Fe                                  | ND       | ND   | 0.01  | 0.02  | 0.03       | Forsterite           |       |       |             |       |      |       |
| lotas cation                        | 5.02     | 5.01 | 1.99  | 3.99  | 2.01       |                      |       |       |             |       |      |       |
|                                     |          |      |       |       |            |                      |       |       |             |       |      |       |
|                                     |          |      |       |       |            |                      |       |       |             |       |      |       |
|                                     |          |      |       |       |            |                      |       |       |             |       |      |       |
|                                     |          |      |       |       |            |                      |       |       |             |       |      |       |
|                                     |          |      |       |       |            |                      |       |       |             |       |      |       |
|                                     |          |      |       |       |            | Anorthite            |       |       |             |       |      |       |
|                                     |          |      |       |       |            |                      |       |       |             |       |      |       |
|                                     |          |      |       |       |            |                      |       |       |             |       |      |       |
|                                     |          |      |       |       |            |                      |       |       |             |       |      |       |
|                                     |          |      |       |       |            |                      |       |       |             |       |      |       |
|                                     |          |      |       |       |            |                      |       |       |             |       |      |       |
|                                     |          |      |       |       |            |                      |       |       |             |       |      |       |
|                                     |          |      |       |       |            | Gehlenite            |       |       |             |       |      |       |
|                                     |          |      |       |       |            |                      |       |       |             |       |      |       |
|                                     |          |      |       |       |            |                      |       |       |             |       |      |       |
|                                     |          |      |       |       |            |                      |       |       |             |       |      |       |
|                                     |          |      |       |       |            |                      |       |       |             |       |      |       |
|                                     |          |      |       |       |            |                      |       |       |             |       |      |       |
|                                     |          |      |       |       |            |                      |       |       |             |       |      |       |
|                                     |          |      |       |       |            | Spinel               |       |       |             |       |      |       |
|                                     |          |      |       |       |            |                      |       |       |             |       |      |       |
|                                     |          |      |       |       |            |                      |       |       |             |       |      |       |
|                                     |          |      |       |       |            |                      |       |       |             |       |      |       |
|                                     |          |      |       |       |            |                      |       |       |             |       |      |       |
|                                     |          |      |       |       |            |                      |       |       |             |       |      |       |
|                                     |          |      |       |       |            |                      |       |       |             |       |      |       |

Oxygen isotopic compositions and chemical compositions of each major phase were measured by secondary ion mass spectrometry (using TITech Cameca ims 1270) and X-ray microanalysis (using JEOL JSM-5310LV scanning electron microscope + Oxford LINK-ISIS energy dispersive X-ray spectrometer), respectively. The analytical methods are described elsewhere<sup>20</sup>. Beam overlap, area fractions of inadequate phases in the sputtering crater.  $\sigma_{\text{mean}}$ , standard deviation of the mean determined by statistical variation of secondary ion intensities corresponding to precisions for a series of measurements. The accuracy ( $2\sigma$ ) of the  $\delta$ -values is estimated to be  $\pm 7.1\text{‰}$  for  $\delta^{17}\text{O}$  and  $\pm 5.7\text{‰}$  for  $\delta^{18}\text{O}$ . Three independent analyses were made at position 6. Concentrations of phenocrysts and mesostasis were determined at or near the sputtering craters by spot electron beam ( $\sim 1\text{ }\mu\text{m}$ ) and by rastering beam ( $\sim 7 \times 7\text{ }\mu\text{m}$ ), respectively. Low totals for the mesostasis are due to the porous structure.  $\text{FeO}^*$ , all Fe calculated as  $\text{FeO}$ .  $\text{Ni}^*$ , all Ni calculated as Ni. ND, not detected ( $<0.1\text{ wt\%}$ ). Stolper's mode, according to ref. 14.



**Figure 2** Oxygen-isotopic composition of minerals in the chondrule-bearing CAI. Error bars of individual analyses denote standard deviations corresponding to precisions for a series of measurements (Table 1). Accuracy (95% confidence), estimated from repeated measurements of standard materials, is denoted at lower right. Abbreviations: TF, terrestrial fractionation line; CCAM, carbonaceous chondrite anhydrous mineral mixing line<sup>26</sup>. All phenocrysts depleted in <sup>16</sup>O are embedded in Ca-Al-rich glassy mesostasis that is <sup>16</sup>O-rich. Because O isotopic compositions in the melilite clast were measured in sequence during the same analytical session, the difference between the two  $\delta$ -values in the melilite is much larger than the analytical precisions (also see Supplementary Fig. 2 and Supplementary Table 2). The O isotopic zoning of the melilite clast (more <sup>16</sup>O-rich towards rim) is probably due to O isotopic diffusion during cooling of the last melting event. The <sup>16</sup>O-rich mesostasis seems to be slightly inhomogeneous in its oxygen isotopes. This can be explained by aqueous alteration in the parent body because of rapid O diffusivity in glass. Despite the alteration, the secondary O isotopic modifications are partial because the Al-rich glass mostly occupies the sputtered areas of SIMS analysis.



**Figure 3** Schematic view of CAI and chondrule formation. The solar nebula originally consisted of <sup>16</sup>O-rich dusts and <sup>16</sup>O-poor gas<sup>26</sup>. CAIs having a <sup>16</sup>O-rich signature formed at the inside of the inner edge of the gas disk. CAIs with <sup>16</sup>O-poor signature and chondrules formed in the gas disk. Expected radial excursions of the X-point<sup>22</sup> are shown as red and blue colours. CAIs having both <sup>16</sup>O-rich and <sup>16</sup>O-poor signatures formed at the fluctuation zone of the inner edge. O isotopic exchange of CAI minerals occurred when the <sup>16</sup>O-rich CAI was embedded in the gas disk.

solar nebula (the X-point)<sup>22</sup>, for compound objects consisting of a CAI core and a chondrule mantle<sup>23,24</sup>, and for potential cyclic variation of the O-isotopic composition in the formed solids<sup>25</sup>.

Chondrules and CAIs were possibly formed by X-ray-flare heating, arising as a result of the time-dependent interaction of a protoplanetary disk with the protosolar magnetosphere<sup>22</sup>. Figure 3 is a schematic view of the CAI- and chondrule-forming region around the inner edge of the protosolar disk. In the steady state, <sup>16</sup>O-rich solids<sup>26</sup> form <sup>16</sup>O-rich CAIs at the proto-Sun side of the X-point where the <sup>16</sup>O-poor gas<sup>26</sup> is ionized and removed by the strong electromagnetic forces<sup>22</sup>. This region therefore results in a very high dust/gas ratio<sup>27</sup>. Evaporation, condensation and melting of dusts by the flare heating form <sup>16</sup>O-rich CAIs. Chondrules form outward of the X-point, where the precursor solids are immersed in <sup>16</sup>O-poor gas. Chondrules therefore tend to have <sup>16</sup>O-depleted O-isotopic compositions. <sup>16</sup>O-poor CAIs<sup>28</sup> would form in the region between the forming regions of <sup>16</sup>O-rich CAIs and of <sup>16</sup>O-poor chondrules. However, the O isotopic variation observed in A5 indicates that <sup>16</sup>O-rich and <sup>16</sup>O-poor materials were gathered into one object, suggesting O isotope oscillation in the nebular regions where CAIs and chondrules formed.

The radial excursions of the X-point occurred repeatedly on relatively short timescales, with cycles ranging from days to thousands of years<sup>22</sup>. Synchronously the dust-to-gas ratio in the fluctuation zone changed, owing to the radial fluctuation of the inner edge of the gas disk. On the other hand, solids in the gas disk gradually migrated towards the proto-Sun on a longer timescale, ranging from thousands to million of years<sup>29</sup>. Therefore <sup>16</sup>O-rich dusts and <sup>16</sup>O-poor gas were mixed in different proportions. If X-ray-flare heating occurred with the fluctuation, O isotopic compositions of dusts were changed by the heating event. Thus <sup>16</sup>O-rich and <sup>16</sup>O-poor materials coexist in this fluctuation zone. This model is consistent with previous O isotopic observations: oxygen isotopic exchange of <sup>16</sup>O-rich CAI minerals with the <sup>16</sup>O-poor nebula gas can occur by heating when the CAIs are immersed in the gas<sup>30</sup>; the X-point fluctuations and the solid migrations allow mechanical mixing between <sup>16</sup>O-poor CAI and <sup>16</sup>O-poor chondrule solids<sup>28</sup>.

This model indicates that CAI-forming and chondrule-forming events are not independent, but overlapped during the evolution of the early Solar System. We predict that chondrules having the same age as CAIs, or having the same <sup>16</sup>O-enrichment as CAIs, will be found in chondrites. □

Received 14 October 2002; accepted 28 April 2003; doi:10.1038/nature01699.

- Wood, J. A. in *Chondrules and the Protoplanetary Disk* (eds Hewins, R. H., Jones, R. H. & Scott, E. R. D.) 55–69 (Cambridge Univ. Press, Cambridge, 1996).
- Tatsumoto, M., Unruh, D. M. & Desborough, G. A. U-Th-Pb and Rb-Sr systematics of Allende and U-Th-Pb systematics of Orgueil. *Geochim. Cosmochim. Acta* **40**, 617–634 (1976).
- Chen, J. H. & Tilton, G. R. Isotopic lead investigations on the Allende carbonaceous chondrite. *Geochim. Cosmochim. Acta* **40**, 635–643 (1976).
- Amelin, Y., Krot, A. N., Hatchee, I. D. & Ulyanov, A. A. Lead isotopic ages of chondrules and calcium-aluminum-rich inclusions. *Science* **297**, 1678–1683 (2002).
- Russell, S. S., Srinivasan, G., Huss, G. R., Wasserburg, G. J. & MacPherson, G. J. Evidence for widespread <sup>26</sup>Al in the solar nebula and constraints for nebula time scales. *Science* **273**, 757–762 (1996).
- Bischoff, A. & Keil, K. Al-rich objects in ordinary chondrites: Related origin of carbonaceous and ordinary chondrites and their constituents. *Geochim. Cosmochim. Acta* **48**, 693–709 (1984).
- Srinivasan, G. & Bischoff, A. Magnesium-aluminum study of hibonites within a chondrule-like object from Sharps (H3). *Meteorit. Planet. Sci.* **32**, A148 (1998).
- Krot, A. N. et al. Mineralogical, aluminum-magnesium, and oxygen-isotopic studies of the relic calcium-aluminum-rich inclusions in chondrules. *Meteorit. Planet. Sci.* **34**, A68–A69 (1999).
- Maruyama, S., Yurimoto, H. & Sueno, S. Oxygen isotope evidence regarding the formation of spinel-bearing chondrules. *Earth Planet. Sci. Lett.* **169**, 165–171 (1999).
- MacPherson, G. J., Wark, D. A. & Armstrong, J. T. in *Meteorites and the Early Solar System* (eds Kerridge, J. F. & Matthews, M. S.) 746–807 (Univ. Arizona Press, Tucson, 1988).
- Noguchi, T. Texture and chemical composition of pyroxenes in chondrules in carbonaceous and unequilibrated ordinary chondrites. *Proc. NIPR Symp. Antarct. Meteorites* **2**, 169–199 (1989).
- Jones, R. H., Saxton, J. M., Lyon, I. C. & Turner, G. Oxygen isotopes in chondrule olivine and isolated olivine grains from the CO3 chondrite, ALHA77307. *Meteorit. Planet. Sci.* **35**, 849–857 (2000).
- Wasson, J. T., Rubin, A. E. & Yurimoto, H. CO chondrule evidence for a drift in the nebular oxygen-isotopic composition. [abstract]. *Meteorit. Planet. Sci.* **35**, A166–A167 (2000).
- Stolper, E. Crystallization sequences of Ca-Al-rich inclusions from Allende: An experimental study. *Geochim. Cosmochim. Acta* **46**, 2159–2180 (1982).

15. Itoh, S., Kojima, H., Yurimoto, H., *Lunar Planet. Sci.* XXXI, 1323 (2000) [CD-ROM].
16. Stolper, E. & Paque, J. M. Crystallization sequences of Ca-Al-rich inclusions from Allende: The effects of cooling rate and maximum temperature. *Geochim. Cosmochim. Acta* **50**, 1785–1806 (1986).
17. Yurimoto, H., Morioka, M. & Nagasawa, H. Diffusion in single-crystals of melilite: I. Oxygen. *Geochim. Cosmochim. Acta* **53**, 2387–2394 (1989).
18. Lasaga, A. C. in *Kinetics and Equilibrium in Mineral Reactions* (ed. Saxena, S. K.) 81–114 (Springer, New York, 1983).
19. Wasson, J. T., Yurimoto, H. & Russell, S. S.  $^{16}\text{O}$ -rich melilite in CO3.0 chondrites; possible formation of common,  $^{16}\text{O}$ -poor melilite by aqueous alteration. *Geochim. Cosmochim. Acta* **65**, 4539–4549 (2001).
20. Kojima, T., Yada, S. & Tomeoka, K. Ca-Al-rich inclusions in three Antarctic CO3 chondrites, Yamato-81020, Yamato-82050 and Yamato-790992: Record of low-temperature alteration. *Proc. NIPR Symp. Antarct. Meteorites* **8**, 79–96 (1995).
21. Itoh, S., Kojima, H. & Yurimoto, H. Petrography and oxygen isotopic compositions in refractory inclusions from CO chondrites. *Geochim. Cosmochim. Acta* (in the press).
22. Shu, F. H., Shang, H., Glassgold, A. E. & Lee, T. X-rays and fluctuating X-winds from protostars. *Science* **277**, 1475–1479 (1997).
23. Shu, F. H., Shang, H., Gounelle, M., Glassgold, A. E. & Lee, T. The origin of chondrules and refractory inclusions in chondritic meteorites. *Astrophys. J.* **548**, 1029–1050 (2001).
24. Gounelle, M. *et al.* Extinct radioactivities and protosolar cosmic rays: Self-shielding and light elements. *Astrophys. J.* **548**, 1051–1070 (2001).
25. Yurimoto, H., Asada, Y. & Hirai, K. Oxygen isotopic composition of fine-grained CAIs and genetic relation to coarse-grained CAIs. *Meteorit. Planet. Sci.* **36**, A230 (2001).
26. Clayton, R. N. Oxygen isotopes in meteorites. *Annu. Rev. Earth Planet Sci.* **21**, 115–149 (1993).
27. Krot, A. N., McKeegan, K. D., Leshin, L. A., MacPherson, G. J. & Scott, E. R. D. Existence of an  $^{16}\text{O}$ -rich gaseous reservoir in the solar nebula. *Science* **295**, 1051–1054 (2002).
28. Krot, A. N. *et al.* Refractory calcium-aluminum-rich inclusions and aluminum-diopside-rich chondrules in the metal-rich chondrites Hammadah al Hamra 237 and Queen Alexandra Range 94411. *Meteorit. Planet. Sci.* **36**, 1189–1216 (2001).
29. Cameron, A. G. W. The first ten million years in the solar nebula. *Meteoritics* **30**, 133–161 (1995).
30. Yurimoto, H., Ito, M. & Nagasawa, H. Oxygen isotope exchange between refractory inclusion in Allende and solar nebula gas. *Science* **282**, 1874–1877 (1998).

**Supplementary Information** accompanies the paper on [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank J. T. Wasson and A. E. Rubin for loaning us the thin section and for discussions. We also thank A. Meibom, A. N. Krot and T. J. Fagan for discussions and for improving the English of this Letter. This work was supported by Monbu-Kagaku-sho grants.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to H.Y. ([yuri@geo.titech.ac.jp](mailto:yuri@geo.titech.ac.jp)).

## Generation of nonclassical photon pairs for scalable quantum communication with atomic ensembles

A. Kuzmich, W. P. Bowen, A. D. Boozer, A. Boca, C. W. Chou, L.-M. Duan & H. J. Kimble

Norman Bridge Laboratory of Physics 12-33, California Institute of Technology, Pasadena, California 91125, USA

Quantum information science attempts to exploit capabilities from the quantum realm to accomplish tasks that are otherwise impossible in the classical domain<sup>1</sup>. Although sufficient conditions have been formulated for the physical resources required to achieve quantum computation and communication<sup>2</sup>, there is a growing understanding of the power of quantum measurement combined with the conditional evolution of quantum states for accomplishing diverse tasks in quantum information science<sup>3–5</sup>. For example, a protocol has recently been developed<sup>6</sup> for the realization of scalable long-distance quantum communication and the distribution of entanglement over quantum networks. Here we report the first enabling step in the realization of this

protocol, namely the observation of quantum correlations for photon pairs generated in the collective emission from an atomic ensemble. The nonclassical character of the fields is demonstrated by the violation of an inequality involving their normalized correlation functions. Compared to previous investigations of non-classical correlations for photon pairs produced in atomic cascades<sup>7</sup> and in parametric down-conversion<sup>8</sup>, our experiment is distinct in that the correlated photons are separated by a programmable time interval (of about 400 nanoseconds in our initial experiments).

The theoretical proposal of ref. 6 (hereafter ‘DLCZ’) is a probabilistic scheme based upon the entanglement of atomic ensembles via detection events of single photons in which the sources are intrinsically indistinguishable, and generates entanglement over long distances via a quantum repeater architecture<sup>9</sup>. The DLCZ scheme, with built-in quantum memory and entanglement purification, is well within the reach of current experiments and accomplishes the same objectives as previous more complex protocols that require as yet unattainable capabilities<sup>9,10</sup>.

In our experiment, we demonstrate a basic primitive integral to the DLCZ scheme. Specifically, an initial ‘write’ pulse of (classical) light is employed to create a state of collective atomic excitation as heralded by photoelectric detection of a first photon 1. After a programmable delay  $\delta t$ , a subsequent ‘read’ pulse interrogates the atomic sample, leading to the emission of a second (delayed) photon 2. The manifestly quantum (or nonclassical) character of the correlations between the initial ‘write’ photon 1 and the subsequent ‘read’ photon 2 is verified by way of the observed violation of a Cauchy–Schwarz inequality for coincidence detection of the (1, 2) fields<sup>7</sup>. Explicitly, we find  $[\tilde{g}_{1,2}^2(\delta t) = (5.45 \pm 0.11)] \neq [\tilde{g}_{1,1}\tilde{g}_{2,2} = (2.97 \pm 0.08)]$ , where  $\tilde{g}_{ij}$  are normalized correlation functions for the fields ( $i, j$ ) and  $\delta t = 405$  ns is the time separation between the (1, 2) emissions. The capabilities realized in our experiment provide an important initial step towards the implementation of the full DLCZ protocol, which would enable the distribution and storage of entanglement among atomic ensembles distributed over a quantum network. Extensions of these capabilities could facilitate scalable long-distance quantum communication<sup>6</sup> and quantum state engineering<sup>11</sup>. For example, by employing spin-polarized samples in optical-dipole or magnetic traps<sup>12</sup>, it should be possible to extend the interval  $\delta t$  to times of several seconds.

Our experiment arises within the context of prior work on spin squeezing<sup>13,14</sup>, and in particular on atomic ensembles where significant progress has been made in the development of methods to exploit collective enhancement of atom–photon interactions provided by optically thick atomic samples<sup>15–20</sup>. Instead of homodyne or heterodyne detection of light as used in spin-squeezing experiments<sup>18–20</sup>, the DLCZ scheme involves photon-counting techniques, which present stringent requirements for broad bandwidth detection and for the suppression of stray light from the atomic ensemble.

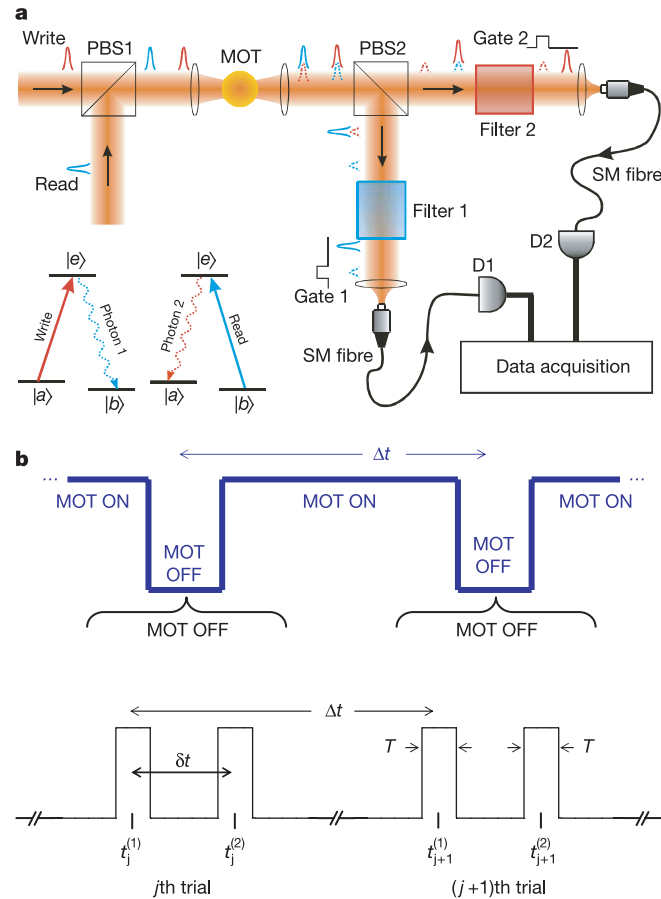
As illustrated in Fig. 1, an optically thick sample of three-level atoms in a lambda-configuration is exploited to produce correlated photons via the following sequence. With atoms initially prepared in state  $|a\rangle$  by optical pumping, a laser pulse from the ‘write’ beam tuned near the  $|a\rangle \rightarrow |e\rangle$  transition illuminates the sample and induces spontaneous Raman scattering to the initially empty level  $|b\rangle$  via the  $|e\rangle \rightarrow |b\rangle$  transition at time  $t^{(1)}$ . The ‘write’ pulse is made sufficiently weak so that the probability to scatter one Raman photon into the preferred forward-propagating mode  $\psi^{(1)}(\mathbf{r}, t)$  is much less than unity for each pulse. Detection of a photon in the mode  $\psi^{(1)}(\mathbf{r}, t)$  produced by the  $|e\rangle \rightarrow |b\rangle$  transition results in a single excitation in the atomic level  $|b\rangle$  distributed across the



sample. In the ideal case, this coherently symmetrized state is<sup>6</sup>:

$$|\Phi_{1A}\rangle \approx \sum_{j=1}^N |a\rangle_1 \dots |b\rangle_j \dots |a\rangle_N \quad (1)$$

Although the initial detection of photon 1 generated by the ‘write’ beam is probabilistic, the detection of photon 1 results in the conditional state  $|\Phi_{1A}\rangle$  with one collective atomic ‘excitation’. This excitation can subsequently be converted into an excitation



**Figure 1** A simplified schematic of the experiment is presented. **a**, Diagram of the apparatus; **b**, the timing sequence for data acquisition. Further details are as follows. **a**, ‘write’ and ‘read’ pulses propagate sequentially into a cloud of cold Cs atoms (MOT), generating pairs of correlated output photons (1,2), with controlled separation  $\delta t$ . Fields with frequency near that of the  $|a\rangle \leftrightarrow |e\rangle$  (or  $|b\rangle \leftrightarrow |e\rangle$ ) transition are coloured red (or blue) here and in Figs 2 and 3. The ‘write’ and ‘read’ pulses have orthogonal polarizations, are combined into a single input at polarizing beam splitter PBS1, and are then focused into the Cs MOT with a waist of approximately  $30\ \mu\text{m}$ . The output fields are split by PBS2, which also serves as the first stage of filtering the ‘write’, ‘read’ beams from the (1,2) fields. For example, field 2 is transmitted by PBS2 to be subsequently registered by detector D2 while the ‘read’ pulse itself is reflected by  $90^\circ$  at PBS2 and then blocked by an acousto-optical modulator that serves as Gate 1. Further filtering is achieved by passing each of the outputs from PBS2 through separate frequency filters, each of which consists of a glass of Cs vapour optically pumped to place atoms into either  $6S_{1/2}, F=3$  or  $F=4$  (ref. 28). The small residual reflected (transmitted) light of the ‘write’ or ‘read’ pulse from PBS2 at frequency  $\omega_{4,4}$  (or  $\omega_{3,4}$ ) passes through a filter cell with atoms in the  $F=4(3)$  level. It is thereby strongly attenuated ( $>10^6$ ), while the accompanying Raman-scattered light as photons 1 (or 2) at frequency  $\omega_{3,4}$  (or  $\omega_{4,4}$ ) is transmitted with high efficiency ( $\approx 80\%$ ). Transmission efficiencies from the MOT to detectors (D1, D2) are both about 30% for light with the spatial shape of the ‘write’ and ‘read’ beams and of the correct polarization. (D1, D2) have overall quantum efficiencies of approximately 50% (photon in to TTL pulse ‘out’). **b**, Gating windows for the joint detection of photons (1,2) are centred at times  $(t_j^{(1)}, t_j^{(2)})$  for the  $j$ th trial of the experiment during intervals when the MOT is ‘OFF’.

of the light field with high probability ‘on demand’ with a specified emission and a programmable pulse shape<sup>6,10,21,22</sup>. To achieve the conversion from atoms to field, a laser pulse from the ‘read’ beam tuned near the  $|b\rangle \rightarrow |e\rangle$  transition illuminates the atomic sample, thereby affecting the transfer  $|b\rangle \rightarrow |a\rangle$  for the sample with the accompanying emission of a second Raman photon 2 on the  $|e\rangle \rightarrow |a\rangle$  transition. For an optically thick atomic sample, photon 2 is emitted with high probability into a specified mode  $\psi^{(2)}(\mathbf{r}, t)$  offset in time by  $t^{(2)} = t^{(1)} + \delta t$ . The spatial and temporal structure of the modes  $\psi^{(1,2)}(\mathbf{r}, t)$  are set by the geometry of the atomic sample and by the shape and timing of the ‘write’ and ‘read’ beams<sup>21</sup>. In our experiment, the modes of the (write, read) beams are spatially mode-matched, with measured visibility greater than 95% for the case of equal frequency and polarization. The time delay  $\delta t$  is limited in principle only by the coherence time between the levels  $|a\rangle$  and  $|b\rangle$ , which can be long in practice.

The atomic sample for our experiment is provided by caesium atoms in a magneto-optical trap (MOT)<sup>12</sup>, where the Cs hyperfine manifolds  $\{|6S_{1/2}, F=4\rangle, |6S_{1/2}, F=3\rangle, |6P_{3/2}, F'=4\rangle\}$  correspond to the levels  $\{|a\rangle, |b\rangle, |e\rangle\}$ , respectively. As illustrated by the timing diagram in Fig. 1, the MOT is chopped from ON to OFF with  $\Delta t = 4\ \mu\text{s}$ . In each cycle there is a ‘dark’ period of duration  $1\ \mu\text{s}$  when all light responsible for trapping and cooling is gated OFF, with less than 0.1% of atoms measured to remain in the  $F=3$  level at this stage. The  $j$ th trial of the protocol for single photon generation is initiated by a ‘write’ pulse which is resonant with the  $6S_{1/2}, F=4 \rightarrow 6P_{3/2}, F'=4$  transition at frequency  $\omega_{4,4}$  and that has duration  $\approx 51\ \text{ns}$  (full-width at half-maximum, FWHM). A critical parameter for the experiment is the resonant optical thickness  $\gamma_{4,4}$  of the atomic sample<sup>21</sup>. We measure  $\gamma_{4,4} \approx 4\text{--}5$  for cw excitation, corresponding to an attenuation of intensity  $\exp(-\gamma_{4,4})$  in propagation through the MOT.

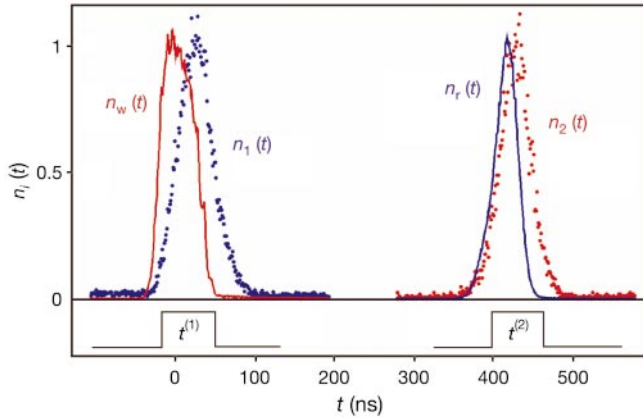
The ‘write’ pulse generates forward-scattered (anti-Stokes) Raman light around frequency  $\omega_{3,4}$  from the  $F'=4$  excited level to the  $F=3$  ground level ( $|e\rangle \rightarrow |b\rangle$ ) that is directed onto a single-photon detector D1. After a variable delay  $\delta t$ , the ‘read’ pulse illuminates the sample, with this pulse tuned to the  $6S_{1/2}, F=3 \rightarrow 6P_{3/2}, F'=4$  transition at frequency  $\omega_{3,4}$  with duration  $\approx 34\ \text{ns}$  (FWHM). Raman (Stokes) light generated by the ‘read’ pulse around frequency  $\omega_{4,4}$  from  $F'=4$  to  $F=4$  ( $|e\rangle \rightarrow |a\rangle$ ) is directed onto a second single-photon detector D2.

By interchanging the frequencies for optical pumping of the filter cells described in Fig. 1, the (write, read) beams can be detected at (D1, D2) instead of the (1, 2) fields. An example of the resulting pulse profiles accumulated over many trials  $\{j\}$  is presented in Fig. 2, where the origin in time is set to coincide with the approximate centre of the ‘write’ pulse, with the ‘read’ pulse following after a delay of  $\approx 415\ \text{ns}$  determined by external control logic.

With the filter cells set to transmit the (1, 2) photons to the (D1, D2) detectors, respectively, we record histograms of the numbers  $(n_1(t), n_2(t))$  of photoelectric events versus time, which are also displayed in Fig. 2. For the data presented here, the intensity of the ‘write’ pulse is kept low ( $\sim 10^3$  photons per pulse), resulting in a time lag for the onset of the  $n_1(t)$  counts in Fig. 2. As discussed in the Supplementary Information, the probability  $p_{\text{write}}^{(1)}$  to generate an anti-Stokes photon 1 within the solid angle of our imaging system is  $p_{\text{write}}^{(1)} \approx 10^{-2}$  per pulse.

The ‘read’ pulse is about 100 times more intense than the ‘write’ pulse, leading to high efficiency  $\xi_{3 \rightarrow 4} \approx 0.6$  for the transfer of population  $|b\rangle \rightarrow |a\rangle$ , with  $p_{\text{read}}^{(2)} \approx \xi_{3 \rightarrow 4} p_{\text{write}}^{(1)}$  for the Stokes photon 2. Examples of the resulting detection events  $n_2(t)$  are shown in Fig. 2. In contrast to the behaviour of  $n_1(t)$ , the intense ‘read’ beam generates  $n_2(t)$  counts promptly. More extensive investigations of the timing characteristics of the emitted fields (1, 2) will be part of our subsequent investigations, including the relationship to electromagnetically induced transparency (EIT)<sup>23,24</sup>.

A virtue of the DLC protocol is its insensitivity to a variety of

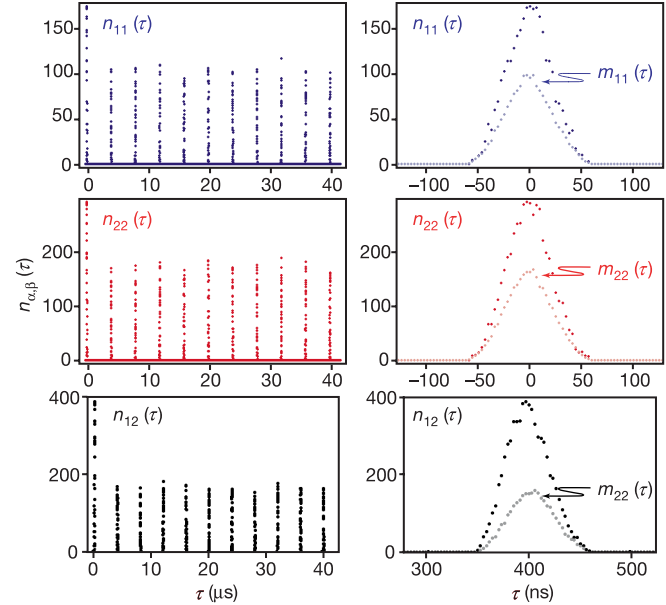


**Figure 2** Normalized singles counts  $n_i(t)$  are shown for the ‘write’, ‘read’ and (1,2) fields. The pulses around  $t = 0$  are from detector D1 for the ‘write’ beam  $n_w(t)$  (solid trace) and for photon 1,  $n_1(t)$  (data points). The pulses around  $t = 410$  ns are from detector D2 for the ‘read’ beam  $n_r(t)$  (solid trace) and for photon 2,  $n_2(t)$  (points). Note that in addition to the symmetrized excitation, each ‘write’ pulse also transfers several hundred atoms into the  $F = 3$  level owing to spontaneous emission from its near-resonant character. However, atoms transferred into  $F = 3$  via spontaneous decay are spatially uncorrelated, so that their contribution to the signal from the ‘read’ channel is strongly suppressed (by roughly the fractional solid angle collected,  $\delta\Omega/4\pi \approx 4 \times 10^{-5}$ ) as compared to the signal from single-atom excitations of the form  $|\Phi_{1,\lambda}\rangle$  (ref. 6).

loss mechanisms, including inefficiencies in transport and detection of the (1, 2) photons. However, in an actual experiment, various non-ideal characteristics of the atom-field interaction (as in our MOT) do lead to deterioration of correlation for the (1, 2) photons (for example, imperfect filtering and/or background fluorescence as described in the caption of Fig. 2 and in the Supplementary Information). Fortunately there exists a well-defined border between the classical and quantum domains for the (1, 2) fields that can be operationally accessed via coincidence detection, as was first demonstrated in the pioneering work of Clauser<sup>7</sup>.

As illustrated in Fig. 1b, electronic pulses from detectors (D1, D2) are separately gated with windows of duration  $T = 60$  ns centred on times  $(t^{(1)}, t^{(2)})$  corresponding to the approximate peaks of the  $(n_1(t), n_2(t))$  pulses shown in Fig. 2. Photoelectric events that fall within the gate windows are directed to a time-interval analyser (TIA) configured in a standard fashion for measurement of photoelectric correlations<sup>25</sup>. For a ‘start’ event from D1 within the interval  $t_i^{(1)} \pm T/2$  for the  $j$ th trial of the experiment, the TIA records the times of ‘stop’ events from D2 within successive intervals  $t_k^{(2)} \pm T/2$ . Over many repetitions of the experiment, we thereby acquire time-resolved coincidences  $n_{1,2}(\tau)$  between the (1, 2) fields, both within the same trial  $k = j$  and for subsequent trials  $k = j + 1, j + 2, \dots$  (that is, a ‘start’ event from trial  $j$  around time  $t_j$  and a ‘stop’ event from trial  $k$  around time  $t_k$ , where  $t_k = t_j + (k - j)\Delta t$  for  $k = j, j + 1, \dots$ ). By a 50%–50% beam splitter, the field 1 can be directed to detectors (D1, D2), and then in turn the field 2 to (D1, D2). We thus also acquire the time-resolved coincidences  $n_{1,1}(\tau)$  and  $n_{2,2}(\tau)$ .

Figure 3 displays an example of data accumulated in this manner for coincidences  $n_{\alpha,\beta}(\tau)$  between the (1,2), (1,1), and (2,2) beams, with successive peaks separated by the time between trials  $\Delta t = 4 \mu\text{s}$ . Note that there is an excess of coincidence counts in each of the initial peaks for joint detections from the same trial ( $\tau < \Delta t$ ) as compared to  $n_{\alpha,\beta}(\tau)$  from different trials ( $\tau > \Delta t$ ). This excess is shown more clearly in the plots in the right column, which expand the time axis from the left column in Fig. 3. Here, data from successive trials  $k = j + 1, \dots, j + 10$  have been offset to  $\tau < \Delta t$  and then averaged for comparison with  $n_{\alpha,\beta}(\tau)$  from the same trial  $j$  by introducing the quantity  $m_{\alpha,\beta}(\tau) = (1/10) \sum_{k=j+1}^{j+10} n_{\alpha,\beta}(\tau + (k - j)\Delta t)$ . As discussed in the Supplementary Information, statistical



**Figure 3** Time-resolved coincidences  $n_{\alpha,\beta}(\tau)$  between the (1,1), (2,2) and (1,2) fields are displayed versus time delay  $\tau$ . Left,  $n_{\alpha,\beta}(\tau)$  is shown over 11 successive repetitions of the experiment. Right, the time axis is expanded to a total duration of 250 ns with  $\tau = 0$  set to the centre of the gating window  $(t^{(1)}, t^{(2)}, t^{(1)})$  for  $(n_{11}, n_{22}, n_{12})$ , respectively. The larger peak  $n_{\alpha,\beta}(\tau)$  corresponds to detection pairs from the same trial  $j$ , while the smaller peak  $m_{\alpha,\beta}(\tau)$  is for pairs from different trials as defined in the text. Typical acquisition parameters are as follows. Detectors (D1, D2) have average count rates of about  $(400 \text{ s}^{-1}, 250 \text{ s}^{-1})$ , respectively, while background counts with no MOT present are about  $100 \text{ s}^{-1}$ . Counts due to the MOT itself (with ‘write’ and ‘read’ beams blocked) are less than  $(10 \text{ s}^{-1}, 20 \text{ s}^{-1})$  for (D1, D2). Dark counts with the inputs to the fibres blocked are less than  $5 \text{ s}^{-1}$ . All these numbers are for the gated-output mode of data acquisition as in Fig. 1 with  $T = 60$  ns.

independence for trials with  $k \neq j$  is enforced by the experimental protocol of reapplying the MOT and repumping beams after each trial.

From the data in Fig. 3, we determine the total number of coincidences  $N_{\alpha,\beta} = \sum_{\{\tau_i\}} n_{\alpha,\beta}(\tau_i)$  with  $(\alpha,\beta) = (1,2)$  obtained by summing over time bins  $\{\tau_i\}$  for detection within the same trial  $j$ , and  $M_{\alpha,\beta} = \sum_{\{\tau_k\}} m_{\alpha,\beta}(\tau_k)$  obtained from ‘start’ and ‘stop’ events from different trials ( $j \neq k$ ). Fields for which the Glauber–Sudarshan phase-space function is well-behaved (that is, classical fields) are constrained by a Cauchy–Schwarz inequality for the various coincidence counts (Supplementary Information and ref. 25), namely:

$$[\tilde{g}_{1,2}(\delta t)]^2 \leq \tilde{g}_{1,1}\tilde{g}_{2,2} \quad (2)$$

where  $\tilde{g}_{1,1} \equiv N_{1,1}/M_{1,1}$ ,  $\tilde{g}_{2,2} \equiv N_{2,2}/M_{2,2}$ ,  $\tilde{g}_{1,2}(\delta t) \equiv N_{1,2}/M_{1,2}$ .

For the data displayed in Fig. 3, we find  $\tilde{g}_{1,1} = (1.739 \pm 0.020)$  and  $\tilde{g}_{2,2} = (1.710 \pm 0.015)$ , in correspondence to the expectation that the (1,2) fields should each exhibit gaussian statistics with  $\tilde{g}_{1,1} = \tilde{g}_{2,2} = 2$  for the protocol of DLCZ in the ideal case, but here degraded by diverse sources of background counts (see Supplementary Information). By contrast, for the cross-correlations of the (1,2) fields, we record  $\tilde{g}_{1,2}(\delta t) = (2.335 \pm 0.014)$ , with  $\delta t = 405$  ns. Hence the inequality of equation (2) for classical fields is strongly violated, namely  $[\tilde{g}_{1,2}^2(\delta t) = 5.45 \pm 0.11] \not\leq [\tilde{g}_{1,1}\tilde{g}_{2,2} = 2.97 \pm 0.08]$ , where all errors indicate the statistical uncertainties. This violation of the Cauchy–Schwarz inequality clearly demonstrates the non-classical character of the correlations between photons (1,2) generated by the (write, read) beams. Moreover, as discussed in more detail in the Supplementary Information, the measured coincidence rates in Fig. 3 explicitly document the cooperative nature of the

emission process. Overall, we estimate that the probability  $p_c^{(q)}$  for coincidence of the (1,2) photons due to collective atomic excitation as described by the state  $|\Phi_{1A}\rangle$  is roughly  $p_c^{(q)} \approx 10^{-4}$  for each trial  $j$ , referenced to the output of the MOT.

The temporal extent of the photon wave packet  $\psi(\mathbf{r}, t)$  for the (1,2) photons is also of some interest. To investigate this issue, we have carried out the experiment with expanded gate windows of duration  $T = 140$  ns that then encompass the entire domains over which counts  $n_1(t)$  and  $n_2(t)$  are observed in Fig. 2. In this case, we record  $\tilde{g}_{1,1} = (1.72 \pm 0.04)$ ,  $\tilde{g}_{2,2} = (1.52 \pm 0.05)$ , and  $\tilde{g}_{1,2}(\delta t) = (2.45 \pm 0.10)$ , now with  $\delta t$  set to be 320 ns. The classical inequality of equation (2) is once again not satisfied;  $[\tilde{g}_{1,2}(\delta t) = 6.00 \pm 0.50] \not\leq [\tilde{g}_{1,1}\tilde{g}_{2,2} = 2.61 \pm 0.11]$ . These results with  $T = 140$  ns also confirm that dead-time effects do not play a significant role in the current experiment.

As described in the Supplementary Information, the violation of the Cauchy–Schwarz inequality of equation (2) in the ideal case can be much larger than we have observed, namely  $[\tilde{g}_{1,2}(\delta t)]^2 / [\tilde{g}_{1,1}\tilde{g}_{2,2}] \approx [(1+p)/(2p)]^2 \gg 1$ , where  $p \ll 1$  is the excitation probability. In our experiment, the size of the violation of the inequality was limited mostly by uncorrelated fluorescence from individual atoms in the atomic sample. This contribution will be made smaller in future experiments by moving to off-resonant excitation, which necessitates higher optical density. There is also a significant limitation due to the presence of the leakage light from the ‘read’ pulse. This classical pulse is only 9 GHz away from the single-photon field 2 of interest, and is filtered by a factor exceeding  $10^{-9}$ . To achieve even stronger violation of the inequality, we must further improve the filtering capability.

Our observations of nonclassical correlations between the (1,2) photons represent the first important step in the realization of the protocol of DLCZ<sup>6</sup> for scalable quantum communication with atomic ensembles, although it is not yet sufficient for realization of the full protocol. Beyond the nonclassical correlations, our experiment also demonstrates successful filtering of the various fields and collective enhancement by the atomic ensemble, all of which are critical for realization of the full quantum repeater protocol. More generally, the capabilities that we have demonstrated should help to enable other advances in the field of quantum information, including implementation of quantum memory<sup>22,26</sup> and fully controllable single-photon sources<sup>27</sup>, which, when combined, help to pave the avenue for realization of universal quantum computation<sup>4</sup>.

**Note added in proof:** We have been made aware of a recently published related paper by van der Wal et al.<sup>29</sup> □

Received 3 March; accepted 8 May 2003; doi:10.1038/nature01714.

- Preskill, J. P. *Quantum Computation* Lecture Notes for Physics 219/Computer Science 219 (<http://www.theory.caltech.edu/people/preskill/ph229/>).
- DiVincenzo, D. P. The physical implementation of quantum computation. *Fortsch. Phys.* **48**, 771–783 (2000).
- Bose, S., Knight, P. L., Plenio, M. B. & Vedral, V. Proposal for teleportation of an atomic state via cavity decay. *Phys. Rev. Lett.* **83**, 5158–5161 (1999).
- Knill, E., Laflamme, R. & Milburn, G. J. A scheme for efficient quantum computation with linear optics. *Nature* **409**, 46–52 (2001).
- Raussendorf, R. & Briegel, H. J. A one-way quantum computer. *Phys. Rev. Lett.* **86**, 5188–5191 (2001).
- Duan, L.-M., Lukin, M., Cirac, J. I. & Zoller, P. Long-distance quantum communication with atomic ensembles and linear optics. *Nature* **414**, 413–418 (2001).
- Clauser, J. F. Experimental distinction between the quantum and classical field-theoretic predictions for the photoelectric effect. *Phys. Rev. D* **9**, 853–860 (1974).
- Mandel, L. Quantum effects in one-photon and two-photon interference. *Rev. Mod. Phys.* **71**(2), S274–S282 (1999).
- Briegel, H.-J., Duer, W., Cirac, J. I. & Zoller, P. Quantum repeaters: The role of imperfect local operations in quantum communication. *Phys. Rev. Lett.* **81**, 5932–5935 (1999).
- Enk, S. J., Cirac, J. I. & Zoller, P. Photonic channels for quantum communication. *Science* **279**, 205–208 (1998).
- Duan, L.-M. Entangling many atomic ensembles with laser manipulation. *Phys. Rev. Lett.* **88**, 170402–170405 (2002).
- Metcalfe, H. J. & van der Straten, P. *Laser Cooling and Trapping* (Springer, 1999).
- Kitagawa, M. & Ueda, M. Nonlinear-interferometric generation of number-phase correlated fermion states. *Phys. Rev. Lett.* **67**, 1852–1854 (1991).
- Wineland, D. J., Bollinger, J. J., Itano, W. M. & Moore, F. L. Spin squeezing and reduced quantum noise

- in spectroscopy. *Phys. Rev. A* **46**, R6797–R6800 (1992).
- Kuzmich, A., Molmer, K. & Polzik, E. S. Spin squeezing in an ensemble of atoms illuminated with squeezed light. *Phys. Rev. Lett.* **79**, 4782–4785 (1997).
- Kuzmich, A., Bigelow, N. P. & Mandel, L. Atomic quantum non-demolition measurements and squeezing. *Europhys. Lett. A* **42**, 481–486 (1998).
- Molmer, K. Twin-correlations in atoms. *Eur. Phys. J. D* **5**, 301–305 (1999).
- Hald, J., Sørensen, J. L., Schori, C. & Polzik, E. S. Spin squeezed atoms: A macroscopic entangled ensemble created by light. *Phys. Rev. Lett.* **83**, 1319–1320 (1999).
- Kuzmich, A., Mandel, L. & Bigelow, N. P. Generation of spin squeezing via continuous quantum non-demolition measurements. *Phys. Rev. Lett.* **85**, 1594–1597 (2000).
- Julsgaard, B., Kozhekin, A. & Polzik, E. S. Experimental long-lived entanglement of two macroscopic objects. *Nature* **413**, 400–403 (2001).
- Duan, L.-M., Cirac, J. I. & Zoller, P. Three-dimensional theory for interaction between atomic ensembles and free-space light. *Phys. Rev. A* **66**, 023818 (2002).
- Fleischhauer, M. & Lukin, M. D. Dark-state polaritons in electromagnetically induced transparency. *Phys. Rev. Lett.* **84**, 5094–5097 (2000).
- Harris, S. E. & Hau, L. V. Nonlinear optics at low light levels. *Phys. Rev. Lett.* **82**, 4611–4614 (1999).
- Zibrov, A. S. et al. Transporting and time reversing light via atomic coherence. *Phys. Rev. Lett.* **88**, 103601 (2002).
- Mandel, L. & Wolf, E. *Optical Coherence and Quantum Optics* (Cambridge Univ. Press, Cambridge, UK, 1995).
- Schori, C., Julsgaard, B., Sørensen, J. L. & Polzik, E. S. Recording quantum properties of light in a long-lived atomic spin state: Towards quantum memory. *Phys. Rev. Lett.* **89**, 057903 (2002).
- Pelton, M. et al. Efficient source of single photons: A single dot in a micropost microcavity. *Phys. Rev. Lett.* **89**, 233602 (2002).
- Alexandrov, E. B., Balabas, M. V., Pasgalev, A. S., Verkhovskii, A. K. & Yakobson, N. N. Double-resonance atomic magnetometers: from gas discharge to laser pumping. *Laser Phys.* **6**, 244–251 (1996).
- van der Wal, C. H. et al. Atomic memory for correlated photon states. *Science*, published online 22 May 2003 (doi:10.1126/science.1085946).

**Supplementary Information** accompanies the paper on [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** H.J.K. gratefully acknowledges interactions with M. D. Lukin about various aspects of the experiment. This work was supported by the National Science Foundation, by the Caltech MURI Center for Quantum Networks, and by the Office of Naval Research.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to H.J.K. (hjkimble@caltech.edu).

## Ultrafast terahertz probes of transient conducting and insulating phases in an electron–hole gas

R. A. Kaindl\*, M. A. Carnahan\*, D. Hägele\*†, R. Löwenich\* & D. S. Chemla\*

\* Department of Physics, University of California at Berkeley, and Materials Sciences Division, E. O. Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA

† Institut für Festkörperphysik, Universität Hannover, Appelstraße 2, 30167 Hannover, Germany

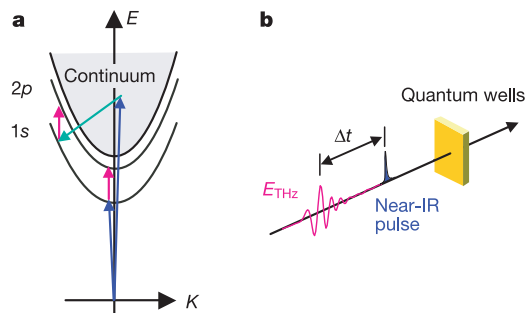
Many-body systems in nature exhibit complexity and self-organization arising from seemingly simple laws. For example, the long-range Coulomb interaction between electrical charges has a simple form, yet is responsible for a plethora of bound states in matter, ranging from the hydrogen atom to complex biochemical structures. Semiconductors form an ideal laboratory for studying many-body interactions of electronic quasiparticles among themselves and with lattice vibrations and light<sup>1–4</sup>. Oppositely charged electron and hole quasiparticles can coexist in an ionized but correlated plasma, or form bound hydrogen-like pairs called excitons<sup>5,6</sup>. The pathways between such states, however, remain elusive in near-visible optical experiments that detect a subset of excitons with vanishing centre-of-mass momenta. In contrast, transitions between internal exciton levels, which occur in the



far-infrared at terahertz ( $10^{12} \text{ s}^{-1}$ ) frequencies<sup>7–9</sup>, are independent of this restriction, suggesting<sup>10</sup> their use as a probe of electron–hole pair dynamics. Here we employ an ultrafast terahertz probe to investigate directly the dynamical interplay of optically-generated excitons and unbound electron–hole pairs in GaAs quantum wells. Our observations reveal an unexpected quasi-instantaneous excitonic enhancement, the formation of insulating excitons on a 100-ps timescale, and the conditions under which excitonic populations prevail.

In semiconductors, optical excitation can promote electron (e) quasiparticles from the highest filled valence band into the empty conduction band separated by an electron–Volt (eV)-sized bandgap. The Coulomb interaction between these depleted electronic states, called holes (h), and the excited electrons can result in the formation of hydrogen-like bound e–h states called excitons<sup>3</sup>. The binding energy of the lowest state is given by  $E_b \propto \mu/(m_0 \epsilon^2) \text{ Ry}$ , where  $\mu$  is the reduced pair effective mass,  $\epsilon$  the dielectric constant,  $m_0$  the bare electron mass, and 1 Ry  $\approx 13.6 \text{ eV}$  the hydrogen atom Rydberg energy. Small reduced masses and large dielectric constants of semiconductors ( $\mu \approx 0.05 m_0$  and  $\epsilon \approx 13$  in prototypical GaAs) result in binding energies of barely a few meV, orders of magnitude below the valence-to-conduction band transition. Hence, the natural scale for electromagnetic transitions between internal exciton levels is the far-infrared, terahertz (THz) spectral range, where 1 THz corresponds to a photon energy of  $\sim 4.1 \text{ meV}$ .

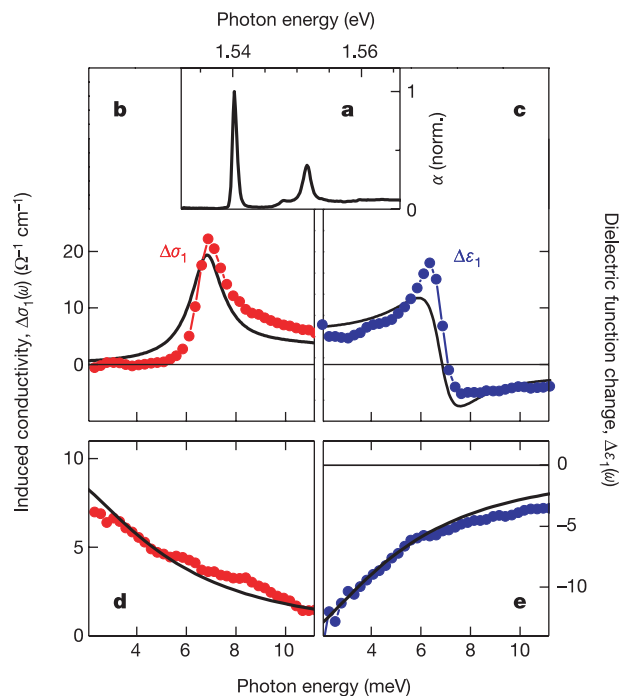
Photoexcitation of even low densities of quasiparticles in condensed matter results in complex many-body states, which arise from the interactions among quasiparticles and with other elementary quanta of vibrations (phonons) or light (photons). In the past, the dynamics of excited semiconductors were often investigated by absorption, photoluminescence, or nonlinear wave mixing experiments close to the bandgap<sup>1,2</sup>. These techniques are based on creation and/or destruction of excitons and—because of the photon’s small momentum—probe e–h states with negligible centre-of-mass momentum  $K$  (Fig. 1a). Absorption in that spectral range measures the generation of e–h pairs, and is consequently only indirectly influenced by carrier and excitonic populations. In contrast, THz transitions can measure populations of excitons or conducting charges already present in the sample—photogenerated or introduced by doping—and at centre-of-mass momenta well outside the optically accessible range<sup>10</sup>. This opens new opportunities for precise spectroscopy of low-energy excitation dynamics.



**Figure 1** Experimental scheme. **a**, Excitation spectrum of e–h pairs with energy  $E$  and centre-of-mass momentum  $K$  in a semiconductor. The 1s and 2s exciton manifolds lie below the continuum of unbound pairs (grey area). Optical pair excitations (blue arrows) can occur only close to  $K \approx 0$  owing to the small photon momentum. Scattering of unbound pairs into exciton states (green arrows) involves large momentum transfer. Transitions between internal exciton states (red arrows) probe excitons at centre-of-mass momenta well outside the optically accessible range. **b**, Carriers are photoexcited in the quantum well by a near-infrared ( $\lambda \approx 800 \text{ nm}$ ) laser pulse of 1 ps duration. A broadband THz pulse of 500 fs duration ( $\lambda \approx 100\text{--}300 \mu\text{m}$ ) arrives with time delay  $\Delta t$  relative to the pump, and detects the THz conductivity and dielectric response of the electrons and holes.

Recently, the build-up of Coulomb screening in an e–h plasma was reported<sup>11</sup>. Furthermore, because time-resolved THz experiments measure both the real and the imaginary parts of the dielectric function, they impose strict conditions on theoretical models<sup>11,12</sup>. These arguments strongly motivate the use of internal exciton transitions to probe exciton dynamics. Early studies observed internal exciton transitions<sup>7–9</sup>, but the dynamics remain largely unexplored. Here, we carry out such experiments, in which changes in a temporally pulsed THz field are detected after it has traversed an optically-excited semiconductor (Fig. 1b). Time-resolved studies under various conditions reveal new and unexpected aspects of such fundamental dynamical processes as formation, population relaxation, and ionization of excitons.

In our experiments, the photoexcited quasiparticles are quantum confined within semiconductor quantum-well nanolayers. We investigate a stack of ten high-quality, undoped 14-nm-wide GaAs wells separated by 10-nm-wide  $\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}$  barriers<sup>13</sup>. The low-temperature near-infrared absorption spectrum (Fig. 2a) is dominated by the absorption line of 1s excitons, built from electrons and heavy holes (HH). With increasing photon energy, other exciton lines appear, followed by the interband e–h pair continuum. The



**Figure 2** Near-infrared and THz properties of GaAs multiple quantum wells. **a**, Near-infrared absorption at the fundamental band edge. With increasing energy, the 1s heavy-hole (HH) exciton absorption line ( $E = 1.540 \text{ eV}$ , full-width at half maximum 0.8 meV, the 2sHH, and the 1s light-hole (LH) absorption lines appear, followed by band-to-band transitions. The HH–LH splitting of  $\sim 11 \text{ meV}$  results from carrier confinement and substrate-induced strain. Photoluminescence spectra (not shown) exhibit a 1-meV line width shifted by  $<0.5 \text{ meV}$  relative to the HH line. After molecular beam epitaxy, the layers were attached to a MgO wafer, and the substrate removed. **b, c**, Induced THz conductivity  $\Delta\sigma_1$  (red) and dielectric function change  $\Delta\epsilon_1$  (blue),  $\Delta t = 5 \text{ ps}$  after resonant excitation of  $\sim 1 \times 10^{10} \text{ cm}^{-2}$  HH excitons at lattice temperature  $T_L = 6 \text{ K}$ . An aperture of 2 mm ensures high transmission of the long-wavelength THz radiation. Black lines: internal exciton response calculated from the dipole moment between the quasi-two-dimensional exciton wavefunctions<sup>16</sup>, with the height scaled to the experimental value. **d, e**, Corresponding response for excitation into the continuum at  $T_L = 300 \text{ K}$ . The black lines show a fit with the Drude model  $\sigma(\omega) = ne^2 d_{\text{QW}}^{-1} m^{-1} (1/\tau - i\omega)^{-1}$ , with  $n = 2.3 \times 10^{10} \text{ cm}^{-2}$ ,  $d_{\text{QW}} = 14 \text{ nm}$ ,  $m = 0.067 m_0$ , and  $1/\tau = 2\pi \cdot 1.15 \text{ THz}$ .  $d_{\text{QW}}$ , quantum well width.

sample is excited with ultrashort near-infrared pump pulses from a 250-kHz Ti:sapphire amplifier spectrally narrowed to 2 meV bandwidth. Picosecond THz probe pulses are generated from the same source by optical rectification, and detected via electro-optic sampling in ZnTe. For each fixed pump–probe time delay  $\Delta t$  between the arrival of pump and probe pulses on the sample (Fig. 1b), we probe the THz field  $E(t)$  transmitted through the unexcited sample and the pump-induced field change  $\Delta E(t)$ . From that, the change  $\Delta\sigma(\omega)$  of the frequency-dependent conductivity in the quantum wells is obtained by Fourier transform of the fields and straightforward electrodynamical relations. Extending earlier single-layer treatments<sup>14</sup>, we account for Fresnel phase-shifts at long wavelengths relevant for our multilayer structure. The complex optical conductivity, expressed as

$$\tilde{\sigma}(\omega) = \sigma_1(\omega) + i \frac{\omega}{4\pi} [1 - \epsilon_1(\omega)]$$

determines the current response  $J(\omega) = \tilde{\sigma}(\omega)E(\omega)$  of the many-body system to the incident transverse electromagnetic field  $E(\omega)$ . Its real part  $\sigma_1(\omega)$  is a measure of the absorbed power density, and the imaginary part is given in terms of the real dielectric function  $\epsilon_1(\omega)$ . Through Kramers–Kronig relations<sup>15</sup>, changes in  $\epsilon_1(\omega)$  are fundamentally linked to a corresponding low-frequency-limit, induced conductivity via

$$\Delta\sigma_1(\omega < 1/\tau) \approx -\frac{1}{2\pi^2} \int_0^\infty \Delta\epsilon_1(\omega') d\omega',$$

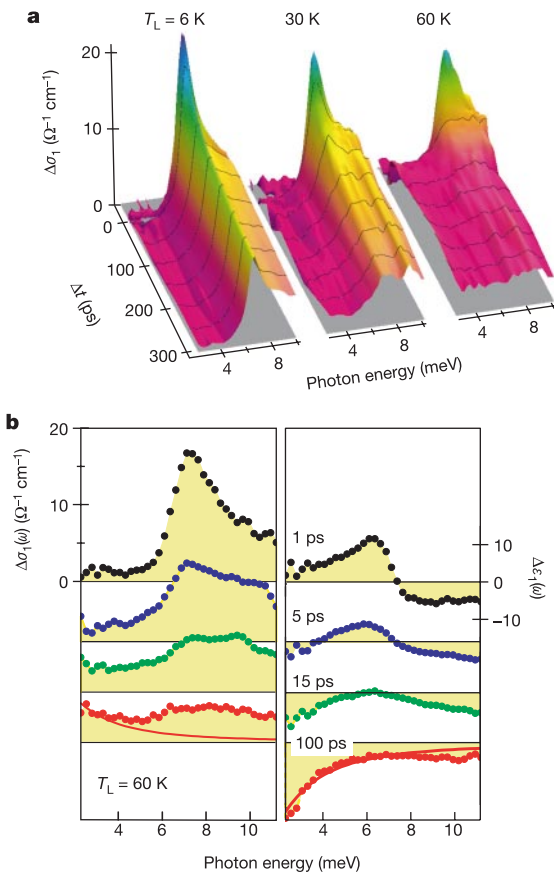
where  $1/\tau$  is the transport scattering rate. Thus, a non-vanishing integral due to an imbalance in  $\Delta\epsilon_1$  signifies a conducting phase. The availability of both parts of  $\tilde{\sigma}(\omega)$ , on an equal footing, is central to understanding the physical state.

Resonant excitation of a density of  $10^{10} \text{ cm}^{-2}$  HH excitons at low lattice temperatures  $T_L = 6 \text{ K}$  leads to the appearance of a strong asymmetric peak in  $\Delta\sigma_1(\omega)$  around  $\omega \approx 7 \text{ meV}$ , as shown in Fig. 2b (dots) 5 ps after excitation. Likewise the corresponding  $\Delta\epsilon_1(\omega)$ , Fig. 2c, indicates the formation of this new low-energy oscillator. This can be explained as follows: initially, the pump creates coherent polarization waves which experience scattering with phonons and defects and within a few picoseconds transform into an incoherent low-energy 1s exciton population (for the near-gap coherent response, see ref. 13). The THz field then probes the internal degrees of freedom of these excitons. Calculations taking into account the quasi-two-dimensional exciton wavefunctions<sup>16</sup> agree closely with this shape (black lines, Fig. 2b, c). Specifically, the peak around 7 meV is attributed to transition from the 1s to the 2p level. Its energy correlates well with reported binding energies<sup>17</sup>. The broad shoulder corresponds to transitions into higher-energy bound and continuum states. The insulating nature of the charge-neutral excitons is apparent from the vanishing low-energy conductivity. A completely different behaviour is seen after excitation of unbound e–h pairs in the continuum (Fig. 2d, e). Here, a high lattice temperature  $T_L = 300 \text{ K}$  is chosen to ensure minimal survival of excitonic correlations, and indeed the response is well-described by the Drude model of dissipative charge motion (black lines). It is important to note that unlike doped semiconductors, the observation of this non-vanishing Drude-like response does not rely on impurities for momentum conservation. Rather, e–h collisions provide efficient scattering channels owing to the large mass difference between these quasiparticles<sup>18</sup>. The conducting nature of this state is evident from both the non-vanishing low-energy conductivity and the all-negative dielectric function change.

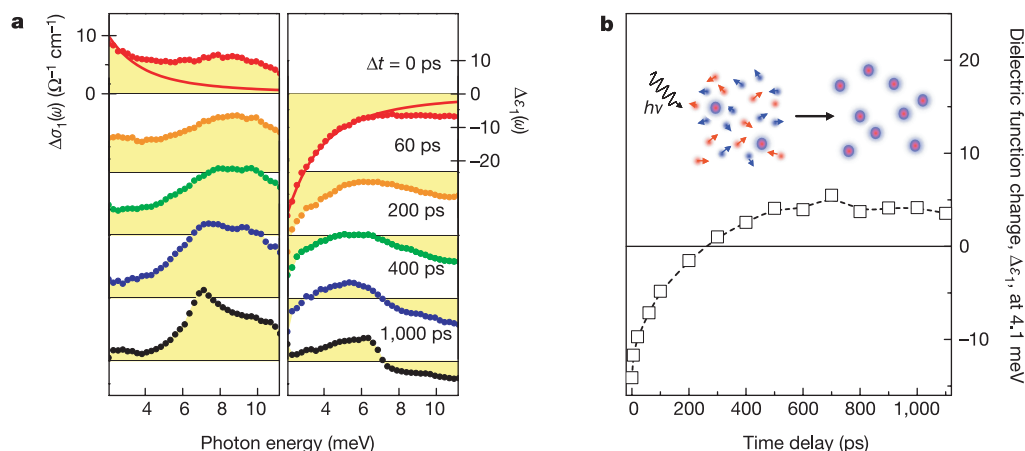
The dynamics after resonant excitation depend strongly on lattice temperature (Fig. 3a). At low temperature ( $T_L = 6 \text{ K}$ ), the conductivity reduces in amplitude but retains its peaked lineshape. Hence its decay follows the recombination of exciton populations within several hundred picoseconds. At higher temperatures ( $T_L = 30 \text{ K}$  and  $60 \text{ K}$ ), the dynamics change considerably. A fast decay of the

peak into a broad response is observed, with a transformation rate that increases strongly with  $T_L$ . Figure 3b shows the induced  $\Delta\sigma_1(\omega)$  and  $\Delta\epsilon_1(\omega)$  spectra for  $T_L = 60 \text{ K}$ . Directly after excitation ( $\Delta t = 1 \text{ ps}$ ) a distinct THz exciton line is observed in  $\Delta\sigma_1(\omega)$ , but with increasing  $\Delta t$  it broadens considerably and eventually decays into the broad response. Correspondingly, the  $\Delta\epsilon_1(\omega)$  profile evolves from the characteristic oscillator dispersion into a negative Drude-like spectrum. Thus the dynamics are governed by ionization (see ref. 19, and references therein) of the cold excitons, via absorption of thermal acoustic and optical phonons. The THz conductivity response of the ionized state ( $\Delta t = 100 \text{ ps}$ , Fig. 3b), however, is not sufficiently described by a Drude model (red line): above 6 meV it exhibits a strong enhancement at the photoionization threshold that indicates the survival of excitonic correlations.

Non-resonant excitation above the bandgap into the unbound e–h continuum leads to a qualitatively different behaviour (Fig. 4a). A broadband THz response is observed at early time delays. The strongly negative  $\Delta\epsilon_1(\omega)$  and large low-frequency conductivity signify a predominantly conducting e–h gas. But surprisingly, a strong excitonic enhancement again exists immediately after excitation ( $\Delta t = 0 \text{ ps}$ ), as evident from the excess conductivity in  $\Delta\sigma_1(\omega)$  around the exciton photoionization threshold. The spectra resemble the time-reversed ionization dynamics of Fig. 3b, although



**Figure 3** Temperature dependence of THz dynamics. **a**, Dynamics of the induced transient conductivity spectra  $\Delta\sigma_1(\omega)$  for three different lattice temperatures of  $T_L$  as indicated, after resonant HH excitation ( $n \approx 1 \times 10^{10} \text{ cm}^{-2}$ ). Colours indicate the magnitude of the induced conductivity change. **b**, Transient spectra for  $T_L = 60 \text{ K}$ . The change of conductivity  $\Delta\sigma_1$  (left panel) and dielectric function  $\Delta\epsilon_1$  (right panel) is shown for several pump–probe time delays. The curves within each panel are equally scaled, but for clarity are shifted in height as indicated by the black line. The red lines ( $\Delta t = 100 \text{ ps}$ ) are the Drude model fits as in Fig. 2d,e, but with  $n = 1.2 \times 10^{10} \text{ cm}^{-2}$  and  $1/\tau = 2\pi \times 0.5 \text{ THz}$ .



**Figure 4** Exciton formation. **a**, Non-resonant excitation with 21 meV excess energy at lattice temperature  $T_L = 6$  K ( $E = 1.561$  eV, photoexcited density  $n \approx 10^{10}$  cm $^{-2}$ ). The change of conductivity (left panel) and dielectric function (right panel) are shown for several time delays  $\Delta t$  between the near-infrared pump and THz probe pulses. Within each panel the curves are equally scaled, but shifted as indicated by the black line. Red

lines ( $\Delta t = 0$  ps): Drude model with  $n = 2 \times 10^{10}$  cm $^{-2}$  and  $1/\tau = 2\pi \times 0.5$  THz. **b**, Time evolution of the induced dielectric function response  $\Delta\epsilon_1$  at  $\omega = 1$  THz (4.1 meV). Inset: schematic evolution of e-h gas dominated by unbound quasiparticles, yet with excitonic correlations, into a pure population of excitons.

with a completely different evolution rate. Features around 7 meV build up on a timescale of hundreds of picoseconds, during which characteristic sharp exciton line shapes emerge and Drude-like contributions abate. Around  $\Delta t = 1,000$  ps, the exciton response is fully restored.

We find two very distinct timescales in exciton formation: one associated with the quasi-instantaneous appearance of a strong excitonic enhancement, and the much slower full transformation from the photoexcited conducting e-h gas to a charge-neutral excitonic phase. Previous theories of exciton formation consider only the extremely dilute limit of independent e-h pairs interacting with phonons<sup>20–22</sup>. This simplified picture cannot account for the complex many-body dynamics observed in real systems and measured in our experiments (Fig. 4). Because excitons are composed of the same states as unbound—yet still interacting—quasiparticles, they cannot be treated as separate species. Within a photoexcited e-h gas, ultrafast Coulomb interactions introduce fundamentally new properties through two-pair and higher correlations. Some of these can be intuitively visualized by processes such as an unbound e-h pair forming an excitonic state after transferring energy and momentum to other carriers (and its reversed process), or redistribution among bound e-h pairs due to collisions with another bound or unbound e-h pair. Such Coulomb interactions naturally explain the observed quasi-instantaneous emergence of an excitonic enhancement around  $\Delta t = 0$  ps in the optical conductivity and its evolution. Our spectra—including intermediate delay times—do not exhibit a sharp exciton structure growing on a broad Drude background, but reveal instead a conglomerate response with time-dependent broadening. This is not explained by centre-of-mass dispersion or phonon scattering that contribute less than  $\sim 1$  meV to broadening<sup>23,24</sup>, but rather reflects the transient nature of the correlations as mediated by Coulomb interactions.

Interactions within the electronic system can, however, only transiently redistribute its internal energy. A full transformation of unbound e-h pairs into excitons requires a phonon bath into which the binding energy and momentum can be permanently released. Photoluminescence measures exciton momentum relaxation to  $K \approx 0$  rather than their formation, and unbound pairs are hard to detect<sup>25–28</sup>. In our experiment, a natural distinction between unbound and excitonic many-particle states arises from the low-energy response that yields a collective measure of charge conductance. The decay of conducting properties observed here is a direct

indicator of the increasingly correlated motion of oppositely charged quasiparticles. This clear distinction arises from the induced dielectric function  $\Delta\epsilon_1(\omega)$ , which even for a broadened oscillator is always positive at frequencies below its resonance, but is generally negative for a Drude-like conductor. Figure 4b illustrates the dynamics of  $\Delta\epsilon_1$  at 1 THz ( $\hbar\omega \approx 4$  meV). The slow monotonic increase from initially negative to positive values is linked to the slowly decreasing contribution from the conducting e-h gas. This gradual process unambiguously shows that exciton formation is characterized by the coexistence of excitons and e-h pairs over several 100 ps.

Binding into excitons eventually forms an insulating quantum state with a fundamental excitation gap equal to the exciton binding energy. In this way, the THz dynamics in Fig. 4 provide evidence for a metal-to-insulator transition<sup>29,30</sup>, which arises here from the decay of a transient conductive state. In the GaAs wells studied here, the fragile nature of the insulating electronic ground state is shown by the marked contrast between its nanosecond persistence after resonant excitation at low temperatures, and the quick, picosecond ionization observed at elevated lattice temperatures. We anticipate that the THz response uncovered here will enable studies of new non-equilibrium states that cannot be investigated directly with near-infrared optics. □

Received 20 December 2002; accepted 15 April 2003; doi:10.1038/nature01676.

- Shah, J. *Ultrafast Spectroscopy of Semiconductors and Semiconductor Nanostructures* (Springer, Berlin, 1999).
- Chemla, D. S. & Shah, J. Many-body and correlation effects in semiconductors. *Nature* **411**, 549–557 (2001).
- Rashba, E. I. *Excitons* (North-Holland, Amsterdam, 1982).
- Perakis, I. Exciton developments. *Nature* **417**, 33–35 (2002).
- Lundstrom, T., Schoenfeld, W., Lee, H. & Petroff, P. M. Exciton storage in semiconductor self-assembled quantum dots. *Science* **286**, 2312–2314 (1999).
- Bayer, M., Stern, O., Hawrylak, P., Fafard, S. & Forchel, A. Hidden symmetries in the energy levels of excitonic 'artificial atoms'. *Nature* **405**, 923–926 (2000).
- Groeneveld, R. H. M. & Grischkowsky, D. Picosecond time-resolved far-infrared experiments on carriers and excitons in GaAs–AlGaAs multiple quantum wells. *J. Opt. Soc. Am. B* **11**, 2502–2507 (1994).
- Timusk, T. Far-infrared absorption study of exciton ionization in germanium. *Phys. Rev. B* **13**, 3511–3514 (1976).
- Černe, J. et al. Terahertz dynamics of excitons in GaAs/AlGaAs quantum wells. *Phys. Rev. Lett.* **77**, 1131–1134 (1996).
- Kira, M., Hoyer, W., Stroucken, T. & Koch, S. W. Exciton formation in semiconductors and the influence of a photonic environment. *Phys. Rev. Lett.* **87**, 176401 (2001).
- Huber, R. et al. How many-particle interactions develop after ultrafast excitation of an electron-hole plasma. *Nature* **414**, 286–289 (2001).
- Beard, M. C., Turner, G. M. & Schmittenmaier, C. A. Transient photoconductivity in GaAs as



- measured by time-resolved terahertz spectroscopy. *Phys. Rev. B* **62**, 15764–15777 (2000).
13. Löwenich, R., Lai, C. W., Hägele, D., Chemla, D. S. & Schäfer, W. Semiconductor polarization dynamics from the coherent to the incoherent regime: Theory and experiment. *Phys. Rev. B* **66**, 045306 (2002).
  14. Nuss, M. C. & Orenstein, J. in *Millimeter and Submillimeter Wave Spectroscopy of Solids* (ed. Grüner, G.) 7–50 (Springer, Berlin, 1998).
  15. Dressel, M. & Grüner, G. *Electrodynamics in Solids* 61–62 (Cambridge Univ. Press, Cambridge, 2002).
  16. Haug, H. & Koch, S. W. *Quantum Theory of the Optical and Electronic Properties of Semiconductors* (World Scientific, Singapore, 1994).
  17. Gerlach, B., Wüsthoff, J., Dzero, M. O. & Smolyanov, M. A. Exciton binding energy in a quantum well. *Phys. Rev. B* **58**, 10568–10577 (1998).
  18. Tzour, N. & Platzman, P. M. High-frequency conductivity of a two-dimensional, two-component electron gas. *Phys. Rev. B* **20**, 4189–4193 (1979).
  19. Wegener, M. *et al.* Femtosecond dynamics of excitonic absorption in the infrared  $\text{In}_x\text{Ga}_{1-x}\text{As}$  quantum wells. *Phys. Rev. B* **39**, 12794–12801 (1989).
  20. Selmann, P. E., Gulia, M., Rossi, F., Molinari, E. & Lugli, P. Coupled free-carrier and exciton relaxation in optically excited semiconductors. *Phys. Rev. B* **54**, 4660–4673 (1996).
  21. Siantidis, K., Axt, V. M. & Kuhn, T. Dynamics of exciton formation for near band-gap excitations. *Phys. Rev. B* **65**, 035303 (2002).
  22. Oh, I. K. Exciton formation assisted by LO phonons in quantum wells. *Phys. Rev. B* **62**, 2045–2050 (2000).
  23. Betz, M. *et al.* Nonlinear optical response of highly energetic excitons in GaAs: Microscopic electrodynamic at semiconductor interfaces. *Phys. Rev. B* **65**, 085314 (2002).
  24. Siarkos, A., Runge, E. & Zimmermann, R. Center of mass properties of the exciton in quantum wells. *Phys. Rev. B* **61**, 10854–10867 (2000).
  25. Damen, T. C. *et al.* Dynamics of exciton formation and relaxation in GaAs quantum wells. *Phys. Rev. B* **42**, 7434–7438 (1990).
  26. Blom, P. W. M., van Hall, P. J., Smit, C., Cuypers, J. P. & Wolter, J. H. Selective exciton formation in thin  $\text{GaAs}/\text{Al}_x\text{Ga}_{1-x}\text{As}$  quantum wells. *Phys. Rev. Lett.* **71**, 3878–3881 (1993).
  27. Deveaud, B., Sermage, B. & Katzer, D. S. Free exciton versus free carrier luminescence in a quantum well. *J. Phys. Colloq. C* **5**, 11–14 (1993).
  28. Kumar, R., Vengurlekar, A. S., Prabhu, S. S., Shah, J. & Pfeiffer, L. N. Picosecond time evolution of free electron-hole pairs into excitons in GaAs quantum wells. *Phys. Rev. B* **54**, 4891–4897 (1996).
  29. Landau, L. & Zeldovich, J. On the relation between the liquid and gaseous states of metals. *Acta Physicochim. USSR XVIII* 2–3, 194–196 (1943).
  30. Mott, N. F. *Metal Insulator Transitions* (Taylor and Francis, London, 1990).

**Acknowledgements** We thank J. Reno for growth of the quantum well samples, and S. L. Chuang for band-structure calculations. We also thank M. Kira, S. W. Koch, M. Woerner, T. Timusk and J. Orenstein for discussions. This work was supported by the Office of Basic Energy Sciences of the US Department of Energy, the Deutsche Forschungsgemeinschaft and the Alexander von Humboldt Foundation.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to R.A.K. (RAKaindl@lbl.gov).

## The ‘zero charge’ partitioning behaviour of noble gases during mantle melting

R. A. Brooker\*, Z. Du†, J. D. Blundy\*, S. P. Kelley‡, N. L. Allan†, B. J. Wood\*, E. M. Chamorro\*§, J.-A. Wartho‡§ & J. A. Purton||

\* CETSEI, Department of Earth Sciences, University of Bristol, Wills Memorial Building, Bristol BS8 1RJ, UK

† School of Chemistry, University of Bristol, Cantock's Close, Bristol BS8 1TS, UK

‡ Department of Earth Sciences, Open University, Walton Hall, Milton Keynes MK7 6AA, UK

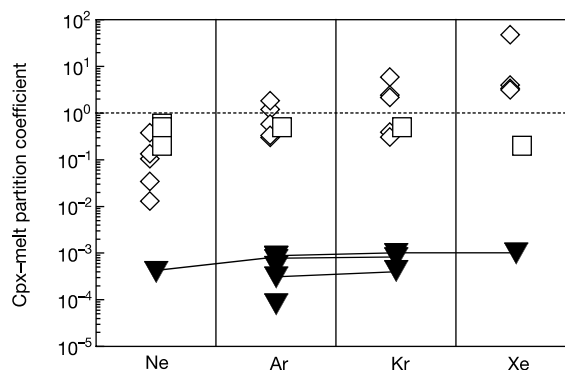
|| CLRC, Daresbury Laboratory, Warrington, Cheshire WA4 4AD, UK

**Noble-gas geochemistry is an important tool for understanding planetary processes from accretion to mantle dynamics and atmospheric formation<sup>1–4</sup>. Central to much of the modelling of such processes is the crystal–melt partitioning of noble gases during mantle melting, magma ascent and near-surface degas-**

**ing<sup>5</sup>. Geochemists have traditionally considered the ‘inert’ noble gases to be extremely incompatible elements, with almost 100 per cent extraction efficiency from the solid phase during melting processes. Previously published experimental data on partitioning between crystalline silicates and melts has, however, suggested that noble gases approach compatible behaviour, and a significant proportion should therefore remain in the mantle during melt extraction<sup>5–8</sup>. Here we present experimental data to show that noble gases are more incompatible than previously demonstrated, but not necessarily to the extent assumed or required by geochemical models. Independent atomistic computer simulations indicate that noble gases can be considered as species of ‘zero charge’ incorporated at crystal lattice sites. Together with the lattice strain model<sup>9,10</sup>, this provides a theoretical framework with which to model noble-gas geochemistry as a function of residual mantle mineralogy.**

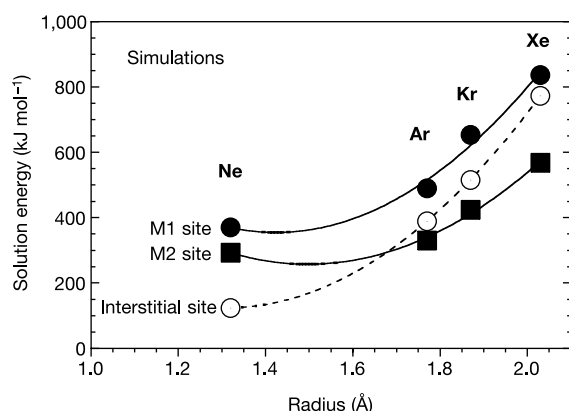
Extreme incompatibility is a basic tenet of most geochemical models for mantle degassing, whereby primordial isotopes such as <sup>3</sup>He and <sup>36</sup>Ar are removed from the mantle and replaced by daughter products, <sup>4</sup>He and <sup>40</sup>Ar, through *in situ* radioactive decay from residual U, Th and K, which are assumed to be less incompatible. Variations in <sup>3</sup>He/<sup>4</sup>He and <sup>36</sup>Ar/<sup>40</sup>Ar ratios in lavas have long been used to argue for a layered mantle in which mid-ocean-ridge basalts (MORB) sample an upper, depleted and degassed layer characterized by low <sup>3</sup>He/<sup>4</sup>He and <sup>36</sup>Ar/<sup>40</sup>Ar ratios, while ocean island basalts (OIB) come from an isolated, undegassed lower (undepleted or ‘primordial’) layer with much higher ratios<sup>1,2,4</sup>. However, this two-layer model has become increasingly difficult to reconcile with independent evidence suggesting whole-mantle convective mixing<sup>11</sup> and tomographic images of subducted slabs in the lower mantle<sup>12</sup>. There are also concerns related to models of the Earth’s radiogenic heat budget<sup>13</sup>, the observation of higher noble-gas abundances in MORB (degassed source) than in OIB (undegassed source)<sup>14</sup> and the distribution of He isotopes between these two ‘source reservoirs’<sup>15</sup>. Consequently, the behaviour of noble gases during mantle melting must be established with some certainty in order to constrain future models of mantle geodynamics.

Our ability to model noble gases during mantle melting is severely compromised by our ignorance of the mechanisms by which noble-gas atoms are incorporated into crystals. There is a widespread view that, unlike trace cations which can clearly enter



**Figure 1** Experimental cpx–melt partition coefficients for noble gases. Values determined for this study using U/LAMP analysis (filled inverted triangles) have errors approximately within the symbols. Data from similar experimental conditions are joined by solid lines. These results are compared to previous experiments using bulk analytical techniques from Hiyagon and Ozima<sup>6</sup> (open squares) and Broadhurst *et al.*<sup>7</sup> (open diamonds). The higher, but parallel, trend observed by Hiyagon and Ozima is the expected consequence of melt contamination, as noted by those authors. Similar results were reproduced, but rejected in this study (see Methods). The systematic fractionation of light over heavy noble gases and *D* values >1 for the Broadhurst *et al.*<sup>7</sup> data appear to be an artefact of adsorbed or trapped gas in their crystals, as discussed in ref. 18.

§ Present addresses: Ecole Normale Supérieure de Lyon, 46 Allée d'Italie, Lyon Cedex 7, 69364 France (E.M.C.); School of Applied Geology, Curtin University, GPO Box U1987, Perth WA 6001, Australia (J.-A.W.).



**Figure 2** Calculated solution energies for noble gases in interstitial positions or lattice sites in jadeite (see Methods). Noble-gas radii in 8-fold coordination are taken from ref. 29.

lattice sites, the ‘inert’ noble gases reside at extended defects, intrinsic defects or grain boundaries, and are therefore not subject to the same thermodynamic controls on crystal–melt partitioning. The situation is not helped by comparison of measurements showing incompatible behaviour for natural samples<sup>16,17</sup>, with contrasting compatible behaviour in early experiments<sup>5–8</sup>. The lack of any theoretical framework for noble-gas incorporation into crystals precludes a critical evaluation of these sparse and disparate crystal–melt partitioning data. In this study we present new experimental clinopyroxene–melt partitioning data for a suite of trace elements including Ne, Ar, Kr and Xe. In addition, we also perform an independent set of atomistic simulations to compare with the experimental results. The data for noble gases are then evaluated along with data for other trace cations to determine if trends observed for each isovalent series can be extrapolated to ‘zero

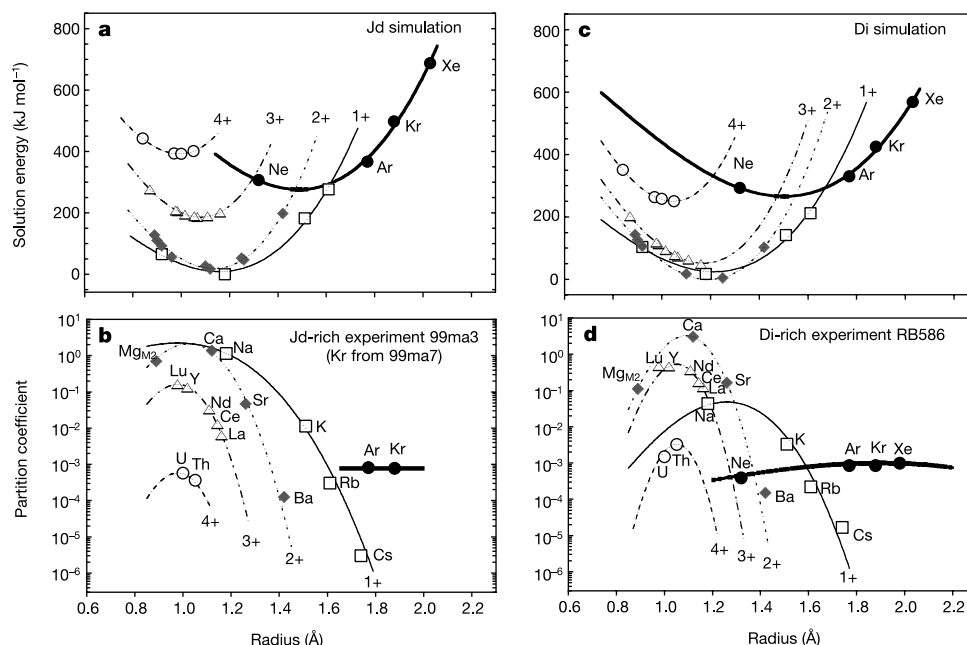
charge’ for the noble gases.

In the experiments, clinopyroxene (cpx) crystals with compositions along the diopside (Di) to jadeite (Jd) solid-solution join were grown from silicate melts doped with a cocktail of trace cations at the 10–1,000 p.p.m. level. Details of experiments are given in Methods, and compositions of crystals and melts are given in Supplementary Information. The concentrations of trace cations in crystals and melt were determined by standard ion-probe techniques, while Ne, Ar, Kr and Xe were measured using an ultraviolet laser ablation microprobe (UVLAMP) technique<sup>18,19</sup>. Nernst partition coefficients ( $D$ ), calculated as the weight concentration ratio of an element in the crystal to that in the melt, are reported in Supplementary Information.

Comparisons of our new UVLAMP results with previously published data for noble gases are shown in Fig. 1. This figure illustrates the relatively low and constant partition coefficient at  $\sim 10^{-3}$ – $10^{-4}$  over a range of cpx composition, pressure (0.1–8.1 GPa) and temperature (1,200–1,665 °C). The discrepancies between the UVLAMP results and older data sets are almost certainly related to the advantages of our new micro-analytical technique in identifying and avoiding fluid or melt inclusions in the crystals, confirming reservations expressed in early studies<sup>6,16,17,20</sup>.

Further support for our lower partition coefficients comes from atomistic computer simulations that follow the approach used in earlier studies of cation incorporation<sup>21</sup>. These calculations evaluate the defect energy, and thus the solution energy associated with the substitution of a noble-gas atom, either on an M1 or M2 site or at an interstitial position in cpx (see Methods).

Calculated solution energies<sup>21</sup> (Fig. 2) indicate that Ar, Kr and Xe substitute at the M2 site in jadeite (and diopside; not shown) rather than the M1 site. The lowest solution energies, which correspond to the highest mineral–melt  $D$  values, involve compensating defects at adjacent M1 sites (for example,  $\text{Ti}^{4+}$  for  $\text{Al}^{3+}$  in jadeite). For Ne, the simulations suggest that an interstitial position is preferred to M1 or



**Figure 3** Computational and experimentally determined Onuma-type<sup>30</sup> diagrams for cpx partitioning. The experimental partition coefficients (**b, d**) and the calculated solution energies for substitution on the M2 site (**a, c**) for each isovalent series of trace elements are plotted against radius<sup>29,31</sup> to show the near-parabolic relationships. Simulated solution energies at M2 (in order of increasing ionic radius) are shown for the noble gases (including Ne at M2); 1+ cations Li, Na, K, Rb; 2+ cations Mg, Co, Fe, Mn, Cd, Ca, Eu, Sr, Ba; 3+ cations Sc, Lu, Yb, Ho, Gd, Eu, Nd, La; and 4+ cations Zr, Ce, U, Th. The curves through each isovalent series for the simulations and experiments have been fitted

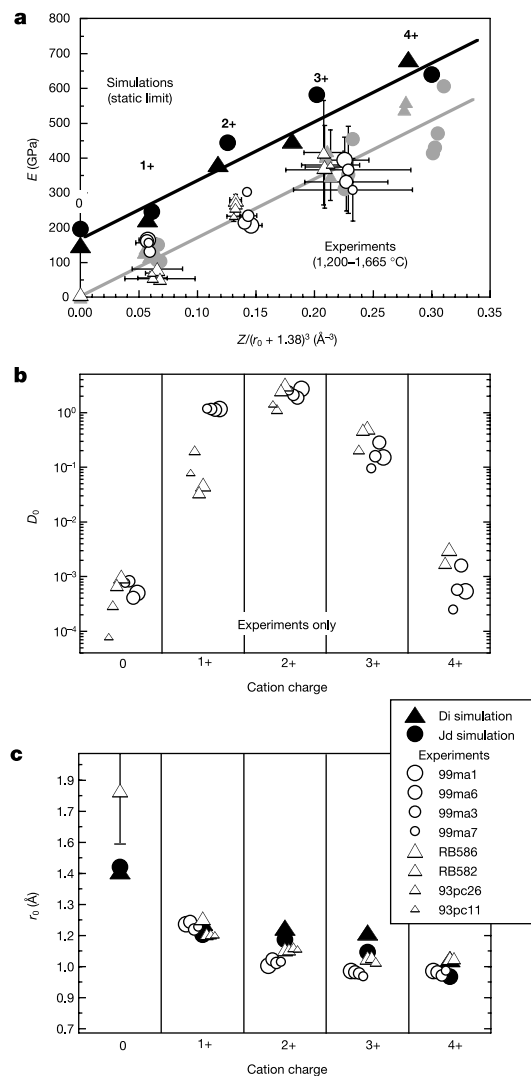
to the lattice strain equations of Brice<sup>9,10,22</sup> (see Methods), except for the limited experimental 4+ data where the fit has been forced through the data points using a proportionate value of  $E$  (see Fig. 4 legend). Similar trends are seen in the other experiments of this study and the derived parameters,  $E$ ,  $r_0$  and  $D_0$ , are presented in Supplementary Information. The curve through the noble gases in **d** is a fitted parabola, but the limited data in **b** are connected by a straight line. Error bars have been omitted for clarity, but are available in Supplementary Information.

M2. In Fig. 3, the partition coefficients for cations that prefer the M2 site are plotted as a function of ionic radius, and compared with those for the noble gases. Data from selected experiments are compared with simulations for pure jadeite and diopside cpx compositions. Both the experimental and simulated data show the expected approximately parabolic dependence on ionic radius for each isovalent series<sup>9,10</sup>. It is also clear that the curvature of the parabola decreases with decreasing cation charge as shown previously<sup>9</sup>. The parabola fitted through the noble-gas data in Fig. 3d is consistent with this trend. The openness of the parabolic fit to experimental data compared to that of the simulations in Fig. 3 is most likely to be a consequence of either the difference in site stiffness due to the much lower temperatures for the calculations (in the static limit) compared with the experiments or some simplifying assumptions made for the simulated melt environment. Both experimental and simulated data suggest that the partition coefficients (or solution energies) for noble gases are approximately the same as for the 4+ cations (U and Th), as predicted in ref. 10. The various lines of agreement between simulations and experiments are remarkable, and argue strongly that lattice site energetics exercise the most important control on partitioning of all species irrespective of charge.

Finally, we consider the simulated and experimental data in terms of the Brice equation<sup>22</sup> and the lattice strain model<sup>9,10</sup>. This equation and model, taken together, predict that, at a given pressure and temperature, the primary thermodynamic controls on partitioning are the elastic and electrostatic work required at a crystal lattice site to accommodate a cation that is different in size and/or charge to the host cation normally resident at that site (see Methods). This approach involves three parameters for each set of isovalent cations entering a given lattice site<sup>9</sup>:  $r_0$ , the optimum radius of the site;  $E$ , the apparent or effective Young's modulus of the site, which controls its tolerance of misfit cations; and  $D_0$ , the partition coefficient for a (fictive) cation with radius  $r_0$ . These parameters can then be modelled as a function of pressure and temperature (and melt composition) as described in Methods. The charge mismatch can be thought of as the 'effective' site charge after substitution, that is, the charge difference between the substituent ion (or species) and the charge at the site in the undoped crystal.

The data in Fig. 4 show that observations made for all three parameters as a function of cation charge can be extended to zero charge for the noble gases. In Fig. 4a,  $E$  decreases linearly as the charge on the substituent decreases. In Fig. 4b,  $D_0$  varies parabolically with charge, with a maximum value at the optimum average site charge (that is, zero effective charge), which is between 2+ and 1+ depending on the diopside:jadeite ratio. The value of  $D_0$  for an effective charge of 2- (that is, the noble gases) is very similar to that for 2+ (that is,  $U^{4+}$  and  $Th^{4+}$ ), as predicted in ref. 10. Accurate determination of  $r_0$  becomes very difficult as the parabola become more open, but a general increase with decreasing charge is apparent in Fig. 4c. The behaviour of  $D_{Ne}$  in Figs 3 and 4 remains inconclusive. The data could be consistent with Ne occupying the M2 site, Ne at interstitial positions (Fig. 2) with partitioning characteristics similar to M2, or Ne may sit on both M2 and the smaller M1 site. The last two options may increase  $D_{Ne}$  and artificially decrease the fitted value of  $E$  for M2 (that is, to give a more open parabola). However, the constraints provided by the heavier noble gases would still produce the general trends observed in Fig. 4.

The consistency of these systematic trends strongly suggests that noble gases enter crystal lattices in just the same way as cations, causing lattice strains due to the mismatch of size and charge. It is this effective charge which gives the 'inert' noble gases their apparent 'crystal chemistry'. It is unlikely that our observations would be followed if noble gases did not enter lattice sites. For example, interstitial incorporation for all the noble gases would produce different trends in  $r_0$  and  $E$  compared with values obtained by extrapolation from those for cations. Another possibility is incorporation at extended defects: this would result in a continuous



**Figure 4** Comparison of lattice strain model parameters. Data derived from both simulations (filled black symbols) and experiments (open symbols) are plotted as a function of charge on the substituent species: **a**,  $E$ , the apparent site Young's modulus; **b**,  $D_0$ , the partition coefficient for a cation with ideal radius  $r_0$ ; and **c**,  $r_0$ , the optimum site radius. Triangles are Di-rich samples, circles Jd-rich. In **a**,  $E$  is plotted against the ratio of charge to site volume ( $d^3$ ), where  $d = r_0 + 1.38 \text{ \AA}$  (the ionic radius of  $O^{2-}$ ; ref. 31). The  $E$  values for the unconstrained fitted parabola through 1+, 2+ and 3+ cation trends and the noble gases in Fig. 3d are shown as open symbols in **a**, error bars reflecting the range of possible fits through the analytical errors on measured partition coefficients. However, it is not possible to constrain a parabola through the two experimental 4+ data points. An alternative approach is to fit the data by using the values of  $E$  for the best-defined parabola (2+), and calculating 1/2, 3/2 or 4/2 times this value for 1+, 3+ and 4+, respectively<sup>9</sup>. These fits are shown as grey symbols in **a**, and there is good agreement between the  $E$  values derived by each method. These proportionate 4+  $E$  values are used to fit the experimental data for 4+ cations in Fig. 3b and d, and in **b** and **c** of this figure. Differences between experimental and simulation data trends reflect the difference in temperature between the two data sets (static limit versus 1,200–1,665 °C) and the simplifying assumptions about melt environment used in the simulations. In **b**,  $D_0$  shows an approximately parabolic dependence on charge<sup>10</sup>. The variation in peak position for different experiments reflects variation in optimum M2-site charge from 1+ (jadeite) to 2+ (diopside). In **c**,  $r_0$  is seen to generally increase with decreasing charge. Parabola shape and peak positions are generally difficult to constrain, as data are biased towards the large radii limb of the parabola. This is further compounded for the noble gases by the low curvature of the parabola, and a representative large error bar has been included in **c**.



decrease in partition coefficient with increasing atomic size, as observed for noble-gas adsorption<sup>6,18</sup>. Incorporation at intrinsic defects is not a viable mechanism for the noble-gas concentration range in our experiments (~0.01–0.5 p.p.m.).

The ability to consider noble gases as ‘zero charge’ cations represents an advance in our capability to model heavy noble gas behaviour during mantle melting. Noble gases need not be viewed as ‘special cases’ with partitioning behaviour dominated by intractable defect-controlled or grain-boundary processes. Comprehensive modelling of noble gases during mantle melting will require further data for other important crystalline phases to derive bulk incompatibilities as a function of residual mantle composition. In terms of radiogenic parents, Ar appears only slightly more incompatible than K in cpx (Fig. 3d), and it remains to be seen how He (possibly not controlled by the M2 site) behaves relative to U and Th. However, partitioning measurements for natural samples and our preliminary experimental results<sup>18</sup> suggest a  $D_{Ar}$  for olivine that is similar to, or higher than, the  $D_{Ar}$  in cpx, whilst K, U and Th would be expected to be relatively more incompatible<sup>23</sup>. If  $D_{He}$  in olivine were also to plot on an open parabola within an order of magnitude of  $D_{Ar}$ , these results would not support the long-standing assumption that noble gases are *de facto* more incompatible than their radioactive parents during mantle melting. Although the curvature of the noble-gas parabola for the M2 site in cpx indicates little fractionation between the noble gases regardless of size, the possibility of interstitial locations for He, or partitioning control by an alternative site (for example, M1 in cpx), could explain the observed fractionation of He (and Ne) from heavier noble gases in some MORB lavas<sup>14</sup>. □

## Methods

### Experiments

Details of piston-cylinder (1.0–3.0 GPa) and multi-anvil experiments (5.6–8.1 GPa) can be found in ref. 19, along with descriptions of the UVLAMP analytical technique. Additional 0.1-GPa experiments were performed in a rapid-quench TZM cold-seal apparatus, with a mixture of He, Ne, Ar, Kr and Xe as the pressure medium. Starting materials for TZM experiments were glasses of Albite:Diopside composition (50:50 for RB582 and 60:40 for RB586), taken above the liquidus to 1,300 °C for 6 h, then cooled to 1,250 and 1,200 °C, respectively, at 0.6 °C h<sup>-1</sup>. The resultant cpx crystals are elongated and usually skeletal with numerous melt tubes along the central axis, but have large enough melt-free areas to allow square ablation pits of 50–100 µm a side. Accidental inclusion of melt in the analysis produces high partition coefficients, and the values used in this study are the average of the lowest 2–3 derived partition coefficients<sup>19</sup>. Note that current UVLAMP measurements for He are not considered accurate enough to be included in this study.

### Simulations

The atomistic calculations allow explicitly for ionic relaxation surrounding the incorporated atom(s). Defect energies were calculated using GULP<sup>24</sup> and the conventional two-region approach with an inner region comprising typically 500 ions. A consistent set of interionic potentials<sup>21</sup> was used. For argon–oxygen interactions, Lennard–Jones potentials were derived by fitting to experimental absorption isotherms<sup>25,26</sup> in three zeolites—silicalite, mordenite and 5A—using grand-canonical Monte Carlo<sup>27</sup>. Potentials for other noble gases were obtained using the usual scaling rules. As a test, the resulting potentials for Xe produced an absorption isotherm for silicalite in excellent agreement with experiment<sup>28</sup>. Lennard–Jones  $\epsilon$  parameters for Ne–O, Ar–O, Kr–O and Xe–O were 60.073, 110, 125.1207 and 147.0658 K respectively;  $\sigma$  values were 2.7491, 3.0566, 3.2991 and 3.4841 Å. All calculations are in the static limit, and as a result are representative of extremely low temperatures. We use the defect energies to calculate solution energies<sup>21</sup>, and we do not determine partition coefficients between solid and melt directly. These classical simulations cannot be extended to He owing to quantum effects.

Solution energies for substitution reactions in jadeite were calculated assuming, as previously<sup>2</sup>, that the environment of the displaced Na<sup>+</sup> or Al<sup>3+</sup> ions in the melt is the same as that in the relevant binary oxide. In Fig. 2, the 1 – effective site charge for the noble-gas atom on the Na(M2) site was compensated by exchanging Ti<sup>4+</sup> for Al<sup>3+</sup> on the M1 site. In Fig. 3, the lowest-energy, plausible charge-compensation mechanism for each valence was chosen as follows. For jadeite; noble gases = Ti(M1), 1 + not required, 2 + = Mg(M2), 3 + = Li(M1), 4 + = Li(M1) + Mg(M1). For diopside; noble gases = Sc(MgM1), 1 + = Lu(MgM1), 2 + not required, 3 + = Na(CaM1), 4 + = 2Na(CaM1).

### Lattice strain model

According to the Blundy and Wood<sup>9</sup> adaptation of the Brice equation<sup>22</sup>, the partition coefficient,  $D_n$ , for an ion with charge  $n +$  and radius  $r_i$  entering crystal lattice site  $M$  can be described in terms of three parameters ( $r_{0(M)}^{n+}$ , the ideal radius of that site for cations with this charge;  $E_M^{n+}$ , the apparent (effective) Young’s modulus of the site; and  $D_{0(M)}^{n+}$ , the ‘strain-free’

partition coefficient for a (fictive) ion with radius  $r_{0(M)}^{n+}$ ) according to the expression:

$$D_i = D_{0(M)}^{n+} \times \exp \left\{ \frac{-4\pi N_A E_M^{n+} \left[ \frac{1}{2} r_{0(M)}^{n+} (r_i - r_{0(M)}^{n+})^2 + \frac{1}{3} (r_i - r_{0(M)}^{n+})^3 \right]}{RT} \right\}$$

where  $N_A$  is the Avogadro constant,  $R$  the gas constant and  $T$  is in K.  $D_{0(M)}^{n+}$  and  $E_M^{n+}$  vary with charge  $n +$  and are pressure, temperature and melt-composition dependent, currently requiring experimental calibration<sup>9</sup>.

Received 15 January; accepted 22 April 2003; doi:10.1038/nature01708.

1. Allègre, C. J., Staudacher, T. & Sarda, P. Rare gas systematics: Formation of the atmosphere, evolution and structure of the Earth’s mantle. *Earth Planet. Sci. Lett.* **81**, 127–150 (1987).
2. Allègre, C. J., Hofmann, A. & O’Nions, K. The argon constraints on mantle structure. *Geophys. Res. Lett.* **23**, 3555–3557 (1996).
3. Harper, L. C. & Jacobsen, S. B. Noble gases and Earth’s accretion. *Science* **273**, 1814–1818 (1996).
4. Turner, G. The outgassing history of the Earth’s atmosphere. *J. Geol. Soc. Lond.* **146**, 147–154 (1989).
5. Broadhurst, C. L., Drake, M. J., Hagee, B. E. & Bernatowicz, T. J. Solubility and partitioning of Ar in anorthite, diopside, forsterite, spinel, and synthetic basaltic liquids. *Geochim. Cosmochim. Acta* **54**, 299–309 (1990).
6. Hiyagon, H. & Ozima, M. Partition of noble gases between olivine and basalt melt. *Geochim. Cosmochim. Acta* **50**, 2045–2057 (1986).
7. Broadhurst, C. L., Drake, M. J., Hagee, B. E. & Bernatowicz, T. J. Solubility and partitioning of Ne, Ar, Kr, and Xe in minerals and synthetic basalt melts. *Geochim. Cosmochim. Acta* **56**, 709–723 (1992).
8. Shibata, T., Takahashi, E. & Ozima, M. In *Noble Gas Geochemistry and Cosmochemistry* (ed. Matsuda, J.-I.) 343–354 (Terra Science, Tokyo, 1994).
9. Blundy, J. D. & Wood, B. J. Prediction of crystal–melt partition coefficients from elastic moduli. *Nature* **372**, 452–454 (1994).
10. Wood, B. J. & Blundy, J. D. The effect of cation charge on crystal–melt partitioning of trace elements. *Earth Planet. Sci. Lett.* **188**, 59–71 (2001).
11. Van Keken, P. F. & Ballentine, C. J. Dynamical models of mantle volatile evolution and the role of phase transitions and temperature-dependent rheology. *J. Geophys. Res.* **104**, 7137–7151 (1999).
12. Van der Hilst, R. D., Widiyantoro, S. & Engdahl, E. R. Evidence for deep mantle circulation from global tomography. *Nature* **386**, 578–584 (1997).
13. O’Nions, R. K. & Oxburgh, E. R. Heat and helium in the Earth. *Nature* **306**, 429–431 (1983).
14. Ozima, M. & Igarashi, G. T. The primordial noble gases in the Earth: A key constraint on Earth evolution models. *Earth Planet. Sci. Lett.* **176**, 219–232 (2000).
15. Anderson, D. L. A statistical test of the two reservoir model for helium. *Earth Planet. Sci. Lett.* **193**, 77–82 (2001).
16. Marty, B. & Lussiez, P. Constraints on rare gas partition coefficients from analysis of olivine–glass from a picritic mid-ocean ridge basalt. *Chem. Geol.* **106**, 1–7 (1993).
17. Valbracht, P. J., Honda, M., Staudigel, H., McDougall, I. & Trost, A. P. In *Noble Gas Geochemistry and Cosmochemistry* (ed. Matsuda, J.-I.) 373–381 (Terra Science, Tokyo, 1994).
18. Brooker, R. A., Wartho, J.-A., Carroll, M. R., Kelley, S. P. & Draper, D. S. Preliminary UVLAMP determinations of argon partition coefficients for olivine and clinopyroxene grown from silicate melts. *Chem. Geol.* **147**, 185–200 (1998).
19. Chamorro, E. M. et al. Ar and K partitioning between clinopyroxene and silicate melt to 8 GPa. *Geochim. Cosmochim. Acta* **66**, 507–519 (2002).
20. Hiyagon, H. Constraints on rare gas partition coefficients from analysis of olivine–glass from a picritic mid-ocean ridge basalt—Comments. *Chem. Geol.* **112**, 119–122 (1994).
21. Allan, N. L., Blundy, J. D., Purton, J. A., Lavrentiev, M. Y. & Wood, B. J. in *EMU Notes in Mineralogy* (ed. Geiger, C. A.) **Vol. 3** 251–302 (Eötvös Univ. Press, Budapest, 2001).
22. Brice, J. C. Some thermodynamic aspects of the growth of strained crystals. *J. Crystal Growth* **28**, 249–253 (1975).
23. Beattie, P. The generation of uranium series disequilibrium by partial melting of spinel peridotite: Constraints from partitioning studies. *Earth Planet. Sci. Lett.* **117**, 379–391 (1993).
24. Gale, J. D. GULP: A computer program for the symmetry-adapted simulation of solids. *J. Chem. Soc. Faraday Trans.* **93**, 629–637 (1997).
25. Watanabe, K., Austin, N. & Stapleton, M. R. Investigation of the air separation properties of zeolite types A, X and Y by Monte Carlo simulation. *Mol. Simulat.* **15**, 197–221 (1995).
26. Macedonia, M. D., Moore, D. D. & Maginn, E. J. Adsorption studies of methane, ethane, and argon in zeolite mordenite: Molecular simulations and experiments. *Langmuir* **16**, 3823–3834 (2000).
27. Frenkel, D. & Smit, B. *Understanding Molecular Simulation* (Academic, San Diego, 1996).
28. Pellenq, R. J. M. & Nicholson, D. Grand ensemble Monte Carlo simulation of simple molecules adsorbed in silicalite-zeolite. *Langmuir* **11**, 1626–1635 (1995).
29. Zhang, Y. & Xu, Z. Atomic radii of noble gas elements in condensed phases. *Am. Mineral.* **80**, 670–675 (1995).
30. Onuma, N., Higuchi, H., Wakita, H. & Nagasawa, H. Trace element partition between two pyroxenes and the host lava. *Earth Planet. Sci. Lett.* **5**, 47–51 (1968).
31. Shannon, R. D. Revised effective radii and systematic studies of interatomic distances in halides and chalcogenides. *Acta Crystallogr. A* **32**, 751–767 (1976).

**Supplementary Information** accompanies the paper on [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank J. Craven and R. Hinton for analytical assistance with the Edinburgh ion-probe; we also thank M. Carroll for originally inspiring the idea of a ‘flat’ parabola, and J. Jones for comments. This work was supported by the NERC, and by research fellowships from the Royal Society, the European Commission and the Leverhulme Trust.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to R.A.B. (r.a.brooker@bristol.ac.uk).

## Pleistocene *Homo sapiens* from Middle Awash, Ethiopia

Tim D. White\*, Berhane Asfaw†, David DeGusta\*, Henry Gilbert\*, Gary D. Richards‡, Gen Suwa§ & F. Clark Howell‡

\* Department of Integrative Biology and Laboratory for Human Evolutionary Studies, Museum of Vertebrate Zoology, University of California, Berkeley, California 94720-3160, USA

† Rift Valley Research Service, P.O. Box 5717, Addis Ababa, Ethiopia

‡ Laboratory for Human Evolutionary Studies, Museum of Vertebrate Zoology, University of California, Berkeley, California 94720-3160, USA

§ The University Museum, The University of Tokyo, Hongo, Bunkyo-ku, Tokyo, 113-0033, Japan

The origin of anatomically modern *Homo sapiens* and the fate of Neanderthals have been fundamental questions in human evolutionary studies for over a century<sup>1–4</sup>. A key barrier to the resolution of these questions has been the lack of substantial and accurately dated African hominid fossils from between 100,000 and 300,000 years ago<sup>5</sup>. Here we describe fossilized hominid crania from Herto, Middle Awash, Ethiopia, that fill this gap and provide crucial evidence on the location, timing and contextual circumstances of the emergence of *Homo sapiens*. Radioisotopically dated to between 160,000 and 154,000 years ago<sup>6</sup>, these new fossils predate classic Neanderthals and lack their derived features. The Herto hominids are morphologically and chronologically intermediate between archaic African fossils and later anatomically modern Late Pleistocene humans. They therefore represent the probable immediate ancestors of anatomically modern humans. Their anatomy and antiquity constitute strong evidence of modern-human emergence in Africa.

The fossilized crania of one immature and two adult hominids were recovered with more fragmentary remains in 1997 from Herto Bouri, a set of localities in the Herto Member of the Bouri Formation in the Middle Awash study area of Ethiopia's Afar depression<sup>6</sup>. These new remains are associated with archaeological assemblages containing elements of both Acheulean and Middle Stone Age technocomplexes<sup>6</sup>. The three crania display evidence of post-mortem mortuary practice<sup>6</sup>.

Here we describe the Herto fossils and compare them with Eurasian Neanderthals, and with earlier and later African fossils, to investigate the emergence of anatomically modern humans. The following descriptions focus on characters relevant to the study of Late Pleistocene hominid affinities. Craniodental dimensions are provided in Supplementary Information. The specimens are illustrated in Figs 1 and 2.

The most complete specimen so far recovered from the Upper Herto Member of the Bouri Formation is an adult cranium from Bouri Vertebrate Paleontology Locality 16 (BOU-VP-16/1). Exposure before discovery led to loss of the left side of the calvarium, but vault distortion is limited to a slight movement of rigid plates; the right temporal process of the zygomatic is displaced about 3 mm posteromedially at the frontozygomatic suture, and the internasal midline is shifted about 2 mm right-laterally. The palate is fairly intact, and the entire right facial skeleton is present. The specimen is fully adult, with patent vault sutures and a heavily worn dentition featuring progressively smaller M1, M2 and M3. Heavily worn premolars bear strong superolateral beveling of the fully exposed dentine occlusal platforms (reminiscent of La Ferrassie 1). The cranium, interpreted here as a male, is generally large and robust, with a cranial capacity estimated by teff seed volume (right side doubled) at about 1,450 cm<sup>3</sup>, at the high end of the modern human range.

The BOU-VP-16/1 cranium is long and high in lateral view

(Fig. 1). The distance between the cranial articular eminence and occlusal plane exceeds that observed in a sample of 2,000 modern human crania (American Indian and Predynastic Egyptian). The Baringo KNM-BK-62 mandible approximates the required ramal height for articulation with the Herto specimen, but this mandible is too small in other dimensions for an appropriate fit. The Herto occipital is strongly flexed, with an occipital angle (103°) that is more acute than that in almost all modern humans<sup>7</sup> and marked by a prominent, massive, rugose external occipital protuberance. There is no occipital bun or suprainiac fossa. The mastoid processes are large and projecting, with a mastoid height (37 mm) much greater than the Neanderthal condition and exceeding all but a few modern humans. The zygomatic root is high relative to the level of the external auditory meatus. The superior margin of the temporal squama is high and arched. The root of the maxillary zygomatic process is centred above the first molar. The zygomatic bone is robust in the infraorbital region. The infraorbital plate is oriented paracoronally and marked by a distinct canine fossa.

In anterior view (Fig. 1) the BOU-VP-16/1 cranium shows a broad upper and lower face, with moderate alveolar prognathism. The malar incisura is deep and bounded laterally by a robust malar tubercle. The broken nasal aperture is bounded inferiorly by a sharp nasal sill and prominent spine. The midface combines a broad interorbital area and tall, narrow nasal bones. The glabellar region is prominent, bilaterally arched, rugose, and projects anteriorly over the superomedial orbital corners. The frontal is moderately bossed and slightly receding, offset from the supraorbital torus by a supratrochlear sulcus. The supraorbital torus is differentiated into halves at the level of the (multiple) supraorbital foramina. The flat lateral portion is extremely broad anteroposteriorly (at zygofrontal suture, 18 mm from orbital rim to temporal line), and forms a superoanteriorly facing trigone. There is an extensive frontal sinus that extends laterally to mid-orbit. The great length of the cranium is evident in superior aspect. Its glabella-to-occipital length (219.5 ± 2 mm) exceeds that found in most other fossil hominids (including Skhul and Qafzeh) and a global sample of over 3,000 modern humans<sup>7</sup>. Prominent temporal lines reach to within 35 mm of the sagittal suture, and parallel the latter over most of the parietal's length. Bi-stephanic breadth is only 96 ± 3 mm, well below the modern human mean despite the specimen's size. The angular torus is not prominent, and is fully within the modern human range.

Despite slight distortion of the cranium evident in posterior view, its profile is clearly 'en-maison', with the greatest interparietal breadth high on the vault. The lambdoidal suture is highly complex, bearing numerous ossicles. In inferior view the palate is deep and broad (75.5 ± 1.5 mm external breadth; near modern human maximum). The foramen magnum is large and anteroposteriorly elongate (45 mm, well above the modern human mean). The digastric grooves are deep and bounded medially by prominent juxtamastoid crests. Relative to most modern humans, the masticatory apparatus is well developed, the dentition large and heavily worn, glenoid fossa broad and deep, and pterygoid plates large and flared.

The second major adult specimen (BOU-VP-16/2) was an even larger adult, as judged by matching parts of its preserved temporal bone. It is represented by portions of temporal, frontal, parietal, zygomatic and occipital. Its occipital is not so angled in sagittal profile as that of BOU-VP-16/1, but bears a more prominent occipital crest. Both vaults are thick (see Supplementary Information). A third adult individual represented by a left parietal fragment (BOU-VP-16/43) shows an extensive squamosal overlap of 24 mm, but might have been slightly smaller overall than the other two adults.

The immature cranium BOU-VP-16/5 was found on the surface after its erosion from an indurated sandstone. It had been shattered into more than 180 small fragments from which the cranial vault

and facial portions were restored. A partial dentition comprises both left deciduous molars as well as unerupted fully formed canine and premolar crowns, and a first molar with wear facets. On the basis of modern human standards, we estimate the individual's age at death as 6–7 years. The cranium is morphologically compatible with the Herto adults. Its vault is pentagonal in posterior profile, and the face shows a clear canine fossa and strong malar incisura. The supraorbitals are poorly developed, with pronounced verticality and frontal bossing. As with the adults, the Herto child exhibits a character complex that is distinctly unlike that of Neanderthals<sup>8</sup>.

The Herto crania are thus not Neanderthals. They exhibit none of the notably derived features that are common to those Eurasian specimens attributed to a Neanderthal lineage<sup>9</sup> represented by a multitude of fossils of successive ages, and culminating in the 'classic' Neanderthals. The Herto hominids are contemporaneous with obvious antecedents of the 'classic' Neanderthals, but do not resemble them. The Herto hominids also have derived characters not seen in *Homo erectus* and in other apparently older African specimens such as Bodo, Saldanha and Kabwe, and so cannot be assigned to those groups.



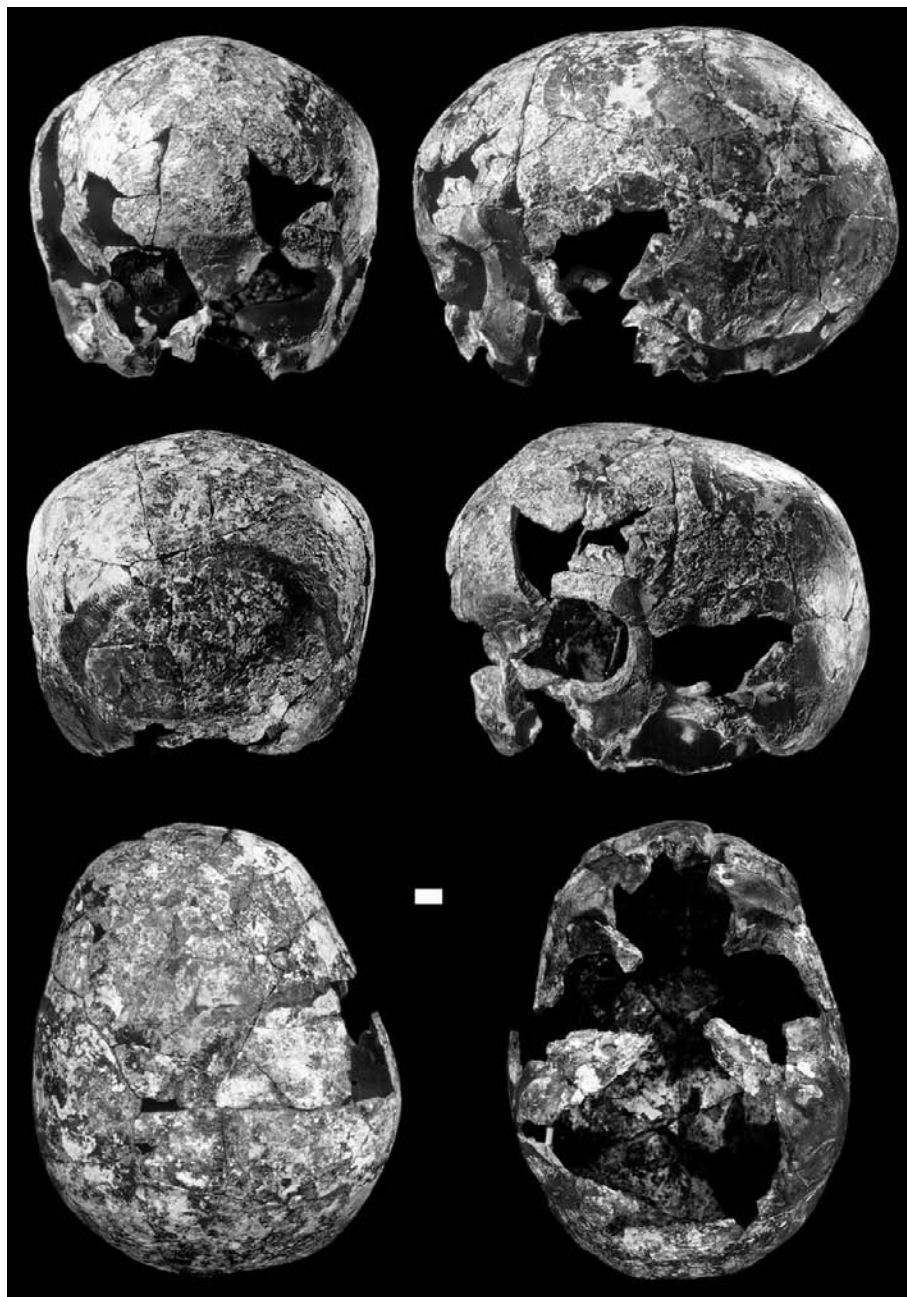
**Figure 1** The Herto BOU-VP-16/1 adult cranium in lateral, frontal, three-quarter, posterior, superior and inferior views. Scale bar, 1 cm.



When BOU-VP-16/1 is compared metrically with a large global sample of modern human crania<sup>7</sup> (Figs 3 and 4), similarities and differences are apparent. The large overall size of BOU-VP-16/1 stands out, aspects of which are described above. Apart from its exceptionally great anterior–posterior length, the cranium exhibits large vault dimensions together with a deep, tall and broad face. However, the orbit and cheekbone dimensions are smaller, and facial projection anterior to the zygomatic is relatively weak. These metric aspects contribute to the comparatively modern gestalt of the face (contrary, for example, to Neanderthals or Kabwe). BOU-VP-16/1 is metrically indistinguishable from anatomically modern *Homo sapiens* (AMHS) in its high cranial vault and relatively large frontal and parietal sagittal dimensions as expressed in their size-standardized variables (by geometric means of 50 variables). Metric

indices of neurocranial globularity and facial retraction have been proposed as diagnostic criteria for AMHS<sup>10</sup>. The former index is estimated at 0.54 in BOU-VP-16/1, within the range suggested to be characteristic of AMHS. Here again, though, the BOU-VP-16/1 parietal bone tends to be less curved, the occipital distinctly flexed, and supraorbitals projecting anteriorly, attesting to its retention of archaic morphology.

Among the global sample of modern humans, the Herto crania, both metrically and non-metrically, lack any derived affinity with modern African crania or with any other modern group, confirming earlier suggestions<sup>11</sup>. Instead, the closest approximations among modern individuals to the overall morphology, size and facial robusticity are found in some Australian and Oceanic individuals, although these are also clearly distinct from the Herto hominids.



**Figure 2** The Herto BOU-VP-16/5 child's cranium in frontal, lateral, posterior, three-quarter, superior and inferior views. Scale bar, 1 cm.

The Herto crania are likewise distinct from Pleistocene representatives of AMHS in some of the features outlined above. In supra-orbital morphology and occipital construction and robusticity, BOU-VP-16/1 is distinguished from the later Klasies and Qafzeh specimens often identified as the earliest AMHS. Other African fossil crania that are possibly temporally intermediate between the early forms (such as Bodo and Kabwe—the ‘early archaic *H. sapiens*’ of Bräuer<sup>12,13</sup>) and AMHS exhibit considerable morphological diversity. The affinities of these specimens (such as Ngaloba, Omo 2, Eliye Springs and Jebel Irhoud—the ‘late archaic *H. sapiens*’ of Bräuer<sup>12,13</sup>) have proved difficult to assess. However, regardless of the particular relationships between these specimens, the general evolutionary position of the Herto sample is clear.

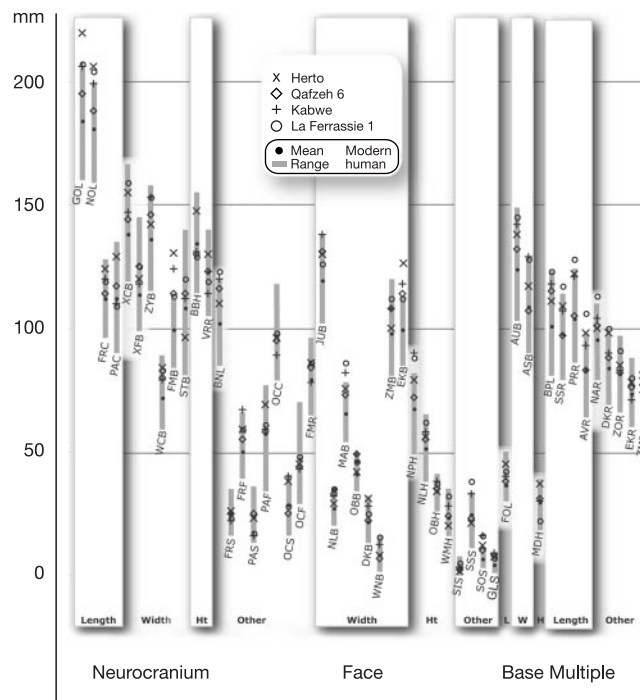
The morphology of the Herto crania falls between the more primitive morphology of the earlier African specimens (such as Bodo and Kabwe) and the more derived morphology of later AMHS (such as Klasies and Qafzeh). The Herto crania are intermediate, metrically and non-metrically, in an African series spanning about 600,000–100,000 years ago, although they are not the only such intermediates in the series. They sample a population that is on the verge of anatomical modernity but not yet fully modern (Fig. 4). This conclusion is supported by comparative anatomical, metric and cladistic considerations, and has profound evolutionary and taxonomic implications.

Some genetic studies<sup>14</sup> have concluded that populations whose contributions quantitatively dominate the modern human gene pool were located in Middle Pleistocene Africa. However, fossil confirmation of these predictions has been lacking. This has prompted some to assert that the sparse African record did not falsify the ‘multiregional’ evolution of AMHS in Europe and the Far East<sup>15–17</sup>. The Herto crania fail to confirm such ‘multiregional’

speculation and conform more closely to most molecular predictions<sup>14,18–20</sup>. They add direct fossil evidence about the anatomy of the populations ancestral to modern humans. The many morphological features shared by the Herto crania and AMHS, to the exclusion of penecontemporaneous Neanderthals, provide additional fossil data excluding Neanderthals from a significant contribution to the ancestry of modern humans.

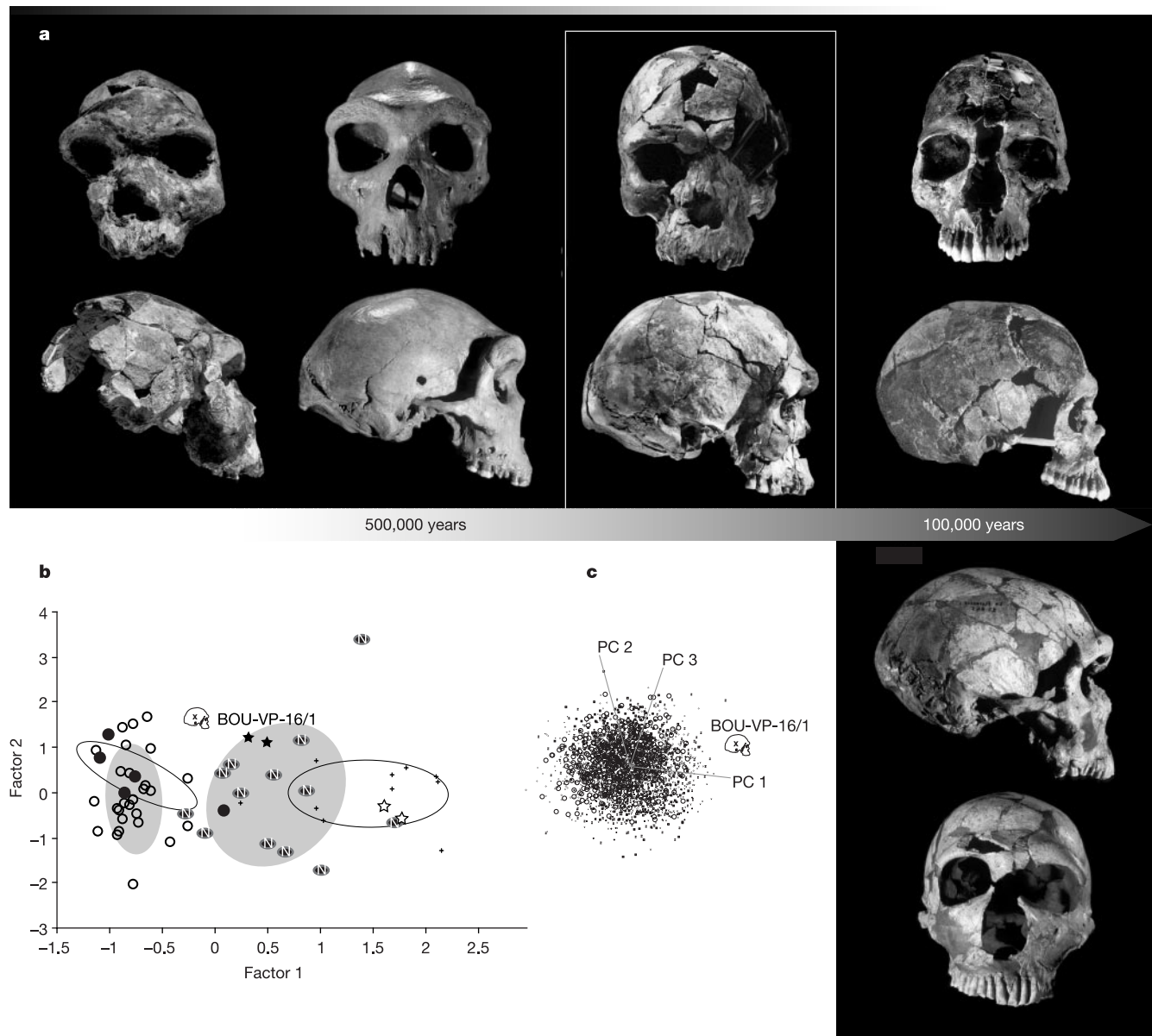
The Herto hominids, although clarifying evolutionary questions, raise taxonomic issues. Widely scattered, often poorly dated, and morphologically diverse Middle and Upper Pleistocene hominid crania from the eastern hemisphere have been assigned to various taxa. In addition to the difficulties inherent in partitioning lineages, several of the available species names are based on inadequate type specimens (such as *H. heidelbergensis*, Schoetensack, 1908; *H. helmei*, Dreyer, 1935; and *H. njarasensis*, Reck and Kohl-Larsen, 1936). Because the Herto hominids are morphologically just beyond the range of variation seen in AMHS, and because they differ from all other known fossil hominids, we recognize them here as *Homo sapiens idaltu*, a new palaeosubspecies of *Homo sapiens* (see Methods). The available evidence from comparative anatomy, multivariate analysis and cladistic considerations suggests that ‘*H. rhodesiensis*’ (Bodo and Kabwe) was ultimately ancestral to *H. sapiens idaltu*, which in turn was ancestral to *Homo sapiens sapiens* (AMHS).

The Middle Awash valley of Ethiopia has now yielded a succession of hominids spanning the past 6 million years<sup>21–26</sup>. Within this study area, and within the genus *Homo*, there exists a chronologically ordered succession of increasingly derived hominids: from Daka (1.0 million years ago) to Bodo (500,000 years ago) to Herto (155,000 years ago). When considered with the evidence from other sites, this



**Figure 3** Univariate comparisons, by anatomical region and dimension, of BOU-VP-16/1 with Qafzeh 6, Kabwe, La Ferrassie, and modern (recent) human males. All comparative data are from Howells<sup>7,27</sup> on original specimens. The listed means are the grand means of the male sample means for the skeletal populations studied by Howells<sup>7,27</sup>. The

measurement abbreviations are as per Howells<sup>7,27</sup>, and are also defined in the Supplementary Information. Symbols: multiplication signs, Herto; diamonds, Qafzeh; plus signs, Kabwe; open circles, La Ferrassie 1; filled circles, means for modern humans; grey bars, range for modern humans. Ht, height.



**Figure 4** **a**, Comparative analysis of the Herto BOU-VP-16/1 adult cranium. **a**, From left to right, in anterior and lateral views, Bodo (The National Museum of Ethiopia, Addis Ababa), Kabwe (The Natural History Museum, London), Herto BOU-VP-16/1 (boxed), Qafzeh 9 (The Rockefeller Museum, Jerusalem), and, inset below, the La Ferrassie Neanderthal (Musée de L'Homme, Paris), all to the same scale, with approximate timeline. **b**, Plot of first two principal component scores, with the position of Herto BOU-VP-16/1 given by the fossil symbol marked 'x'. *Homo erectus*, includes KNM-ER 3733 and KNM-ER 3883 (open stars), and Sangiran, Ngandong and Zhoukoudian crania (plus signs). 'Neanderthals' (Amud, Atapuerca, Gibraltar, La Ferrassie, La Chapelle, La Quina, Monte Circeo, Petralona, Saccopastore, Shanidar, Steinheim and Tabun crania) are shown by circled letter N, Omo 2 and Kabwe by filled stars, and fossil AMHS (Qafzeh 6, 9, Skhul 5, Cro-Magnon and Predmostí 3 crania) by filled circles. Population means of 28 male modern human samples (shown by open circles) were taken from Howells<sup>27</sup> and included in the principal-components analysis (PCA). This was done to show the plot position of modern humans and the degree of inter-population variation. The ovals (line and shade) represent the 1 s.d. dispersion areas of the fossil AMHS sample, the Neanderthal sample,

the Asian *H. erectus* sample, and the modern human population means. The measurements used for this analysis are those of the cranial vault (GOL, NOL, XCB, XFB, AUB, ASB, FRC and PAC; see Supplementary Information for meanings of abbreviations). The raw measurements were standardized for size by the geometric mean of all eight variables. PCA was performed on the size-standardized variables so as to describe shape. The limited number of measurements is a necessary limitation when including fossil specimens. This graph thus illustrates the phenetic affinities as reflected in a limited part of the anatomy. The comparative data are from Arsuaga *et al.*<sup>9</sup> and Howells<sup>27</sup>. **c**, Plot of the first three principal components of the complete Howells<sup>27</sup> data set of 3,024 modern (recent) human individuals, plus Herto BOU-VP-16/1 (the fossil symbol marked 'x'). The principal components were generated from the natural-log-transformed data. This and other results of multivariate analyses demonstrate the phenetic distinctiveness of the Herto hominids relative to modern human crania. However, finer details of the relative position of Herto hominids in multidimensional morphometric space are difficult to interpret because of the necessary inclusion of estimated measurements. The measurements used are given in the Supplementary Information.



shows that modern human morphology emerged in Africa long before the Neanderthals vanished from Eurasia. □

## Methods

Order Primates L., 1758  
Suborder Anthropoidea Mivart, 1864  
Superfamily Hominoidea Gray, 1825  
Family Hominidae Gray, 1825  
*Homo sapiens idaltu* subsp. nov.

**Etymology.** The subspecies name 'idaltu' is taken from the Afar language. It means 'elder'.  
**Holotype.** BOU-VP-16/1 (Fig. 1), an adult cranium with partial dentition. Holotype and referred material are housed at the National Museum of Ethiopia, Addis Ababa. Holotype from Bouri Vertebrate Paleontology Locality 16 (BOU-VP 16); differentially corrected GPS coordinates: 10° 15.5484' N and 40° 33.3834' E.

**Referred material.** BOU-VP-16/2 cranial fragments; BOU-VP-16/3 parietal fragment; BOU-VP-16/4 parietal fragment; BOU-VP-16/5 child's cranium; BOU-VP-16/6 R. upper molar; BOU-VP-16/7 parietal fragment, BOU-VP-16/18 parietal fragments; BOU-VP-16/42 upper premolar, BOU-VP-16/43 parietal fragment.

**Stratigraphy and age.** Bouri Formation, Upper Herto Member. Dated by <sup>40</sup>Ar/<sup>39</sup>Ar to between 160,000 and 154,000 years ago (ref. 6).

**Diagnosis.** On the limited available evidence, a subspecies of *Homo sapiens* distinguished from Holocene anatomically modern humans (*Homo sapiens sapiens*) by greater craniofacial robusticity, greater anterior–posterior cranial length, and large glenoid-to-occlusal plane distance. *Homo sapiens idaltu* is distinguished from the holotype of *Homo rhodesiensis* (Woodward, 1921) by a larger cranial capacity, a more vertical frontal with smaller face, and more marked midfacial topography (for example, canine fossa). We consider the holotypes of *H. helmei* and *H. njarasensis* too fragmentary for appropriate comparisons.

Received 21 November 2002; accepted 14 April 2003; doi:10.1038/nature01669.

- Mbua, E. N. *Patterns of Middle Pleistocene Hominid Evolution in Africa* Dissertation, Univ. Hamburg, (2001).
- Stringer, C. B. Modern human origins: progress and prospects. *Phil. Trans. R. Soc. Lond. B* **357**, 563–579 (2002).
- Howell, F. C. Paleo-demes, species clades, and extinctions in the Pleistocene hominin record. *J. Anthropol. Res.* **55**, 191–243 (1999).
- Stringer, C. B. in *The Age of the Earth: From 4004 BC to AD 2002* (eds Lewis, C. L. E. & Kneel, S. J.) 265–274, Geological Society of London, Spec. Publ. 190 (2001).
- Rightmire, G. P. in *Humanity from African Naisance to Coming Millennia* (eds Tobias, P. V., Raath, M. A., Moggi-Cecchi, J. & Doyle, G. A.) 231–236 (Firenze Univ. Press, 2001).
- Clark, J. D. *et al.* Stratigraphic, chronological and behavioural contexts of Pleistocene *Homo sapiens* from Middle Awash, Ethiopia. *Nature* **423**, 747–752 (2003).
- Howells, W. W. Cranial variation in man. *Pap. Peabody Mus. Archaeol. Ethnol.* **67**, 1–259 (1973).
- Tillier, A.-M. in *The Evolution and Dispersal of Modern Humans in Asia* (eds Akasawa, T., Aoki, K. & Kimura, T.) 15–28 (Hokusen-sha, Tokyo, 1992).
- Arsuaga, J. L., Martínez, I., Gracia, A. & Lorenzo, C. The Sima de los Huesos crania (Sierra de Atapuerca Spain): A comparative study. *J. Hum. Evol.* **33**, 219–281 (1997).
- Lieberman, D. E., McBratney, B. M. & Krovits, G. The evolution and development of cranial form in *Homo sapiens*. *Proc. Natl Acad. Sci. USA* **99**, 1134–1139 (2002).
- Sanich, V. M. in *Conceptual Issues in Modern Human Origins Research* (eds Clark, G. A. & Willermet, C. M.) 392–412 (Aldine, New York, 1997).
- Bräuer, G. in *Humanity from African Naisance to Coming Millennia* (eds Tobias, P. V., Raath, M. A., Moggi-Cecchi, J. & Doyle, G. A.) 191–197 (Firenze Univ. Press, 2001).
- Bräuer, G., (eds Mellars, P. & Stringer, C. B.) 123–154 (Edinburgh Univ. Press, 1989).
- Cann, R. Genetic clues to dispersal in human populations: Retracing the past from the present. *Science* **291**, 1742–1748 (2001).
- Wolpoff, M. & Caspari, R. in *Conceptual Issues in Modern Human Origins Research* (eds Clark, G. A. & Willermet, C. M.) 28–44 (Aldine, New York, 1997).
- Hawks, J. D. & Wolpoff, M. H. The four faces of Eve: Hypothesis compatibility and human origins. *Quat. Int.* **75**, 41–50 (2001).
- Wolpoff, M. H., Hawks, J., Frayer, D. & Hunley, K. Modern human ancestry at the peripheries: A test of the replacement theory. *Science* **291**, 293–297 (2001).
- Templeton, A. R. Out of Africa again and again. *Nature* **416**, 45–51 (2002).
- Underhill, P. A. *et al.* The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Ann. Hum. Genet.* **65**, 43–62 (2001).
- Ingman, M., Kaessmann, H., Pääbo, S. & Gyllenstein, U. Mitochondrial genome variation and the origin of modern humans. *Nature* **408**, 703–713 (2000).
- Haile-Selassie, Y. Late Miocene hominids from the Middle Awash, Ethiopia. *Nature* **412**, 178–181 (2001).
- White, T. D., Suwa, G. & Asfaw, B. *Australopithecus ramidus*, a new species of hominid from Aramis, Ethiopia. *Nature* **371**, 306–312 (1994).
- Lovejoy, C. O., Meindl, R. S., Ohman, J. C., Heiple, K. G. & White, T. D. The Maka femur and its bearing on the antiquity of human walking: Applying contemporary concepts of morphogenesis to the human fossil record. *Am. J. Phys. Anthropol.* **119**, 97–133 (2002).
- White, T. D. *et al.* New discoveries of *Australopithecus* at Maka, Ethiopia. *Nature* **366**, 261–265 (1993).
- Asfaw, B. *et al.* *Australopithecus garhi*: A new species of early hominid from Ethiopia. *Science* **284**, 629–635 (1999).
- Asfaw, B. *et al.* Remains of *Homo erectus* from Bouri, Middle Awash, Ethiopia. *Nature* **416**, 317–320 (2002).
- Howells, W. W. Skull shapes and the map. Craniometric analyses in the dispersion of modern *Homo*. *Pap. Peabody Mus. Archaeol. Ethnol.* **79**, 1–189 (1989).

**Supplementary Information** accompanies the paper on [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank A. Almquist, A. Asfaw, M. Asnake, T. Assebework, D. Brill, J. D. Clark, J. DeHeinzelin, A. Getty, Y. Haile-Selassie, B. Latimer, C. Pehlevan, K. Schick, S. Simpson, P. Snow and Y. Zeleka for fieldwork and analytical studies; J. L. Arsuaga, A. Gracia and N. Garcia for comparative metric data; C. O. Lovejoy for review; and F. Bibi, D. Brill, R. Cann, Y. Haile-Selassie, L. Hlusko, L. Jellema, R. Klein, J. Matternes and R. Paul for assistance. We thank the Ministry of Youth, Sports and Culture, the Authority for Research and Conservation of the Cultural Heritage, and the National Museum of Ethiopia for permissions; the Afar Regional Government and the Afar people of the Middle Awash, particularly the Bouri–Modaitu community and H. Elema; and many other individuals for contributing to our efforts. This research was supported by the NSF (US) and the Japan Society for the Promotion of Science.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to T.W. (timwhite@socrates.berkeley.edu).

## Stratigraphic, chronological and behavioural contexts of Pleistocene *Homo sapiens* from Middle Awash, Ethiopia

J. Desmond Clark\*†, Yonas Beyene‡, Giday WoldeGabriel§, William K. Hart||, Paul R. Renne¶, Henry Gilbert#, Alban Defleur☆, Gen Suwa\*\*, Shigehiro Katoh††, Kenneth R. Ludwig‡‡, Jean-Renaud Boissérie§§, Berhane Asfaw||| & Tim D. White†

\* Department of Anthropology, The University of California, Berkeley, California 94720, USA

‡ Department of Anthropology and Archaeology, Authority for Research and Conservation of Cultural Heritage, P.O. Box 6686, Addis Ababa, Ethiopia  
§ EES-6/MS D462, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA

|| Department of Geology, Miami University, Oxford, Ohio 45056, USA

¶ Berkeley Geochronology Center, 2455 Ridge Road, Berkeley, California 94709, USA, and Department of Earth and Planetary Science, University of California, Berkeley, California 94720, USA

# Department of Integrative Biology and Laboratory for Human Evolutionary Studies, Museum of Vertebrate Zoology, University of California, Berkeley, California 94720, USA

☆ CNRS UMR 6569 du CNRS, Laboratoire d'Anthropologie, Faculté de Médecine, Secteur Nord, Blvd. Pierre Dramart, 13916 Marseille Cedex 20, France

\*\* The University Museum, The University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

†† Hyogo Museum of Nature and Human Activities, Yayoigaoka, Sanda 669-1546, Japan

‡‡ Berkeley Geochronology Center, 2455 Ridge Road, Berkeley, California 94709, USA

§§ CNRS UMR 6046, Université de Poitiers, France

||| Rift Valley Research Service, P.O. Box 5717, Addis Ababa, Ethiopia

† Deceased

Clarifying the geographic, environmental and behavioural contexts in which the emergence of anatomically modern *Homo sapiens* occurred has proved difficult, particularly because Africa lacked adequate geochronological, palaeontological and archaeological evidence. The discovery of anatomically modern *Homo sapiens* fossils at Herto, Ethiopia<sup>1</sup>, changes this. Here we report on stratigraphically associated Late Middle Pleistocene artefacts and fossils from fluvial and lake margin sandstones of the Upper Herto Member of the Bouri Formation, Middle Awash, Afar Rift, Ethiopia. The fossils and artefacts are dated between 160,000 and 154,000 years ago by precise age determinations using the <sup>40</sup>Ar/<sup>39</sup>Ar method. The archaeological assemblages contain elements of both Acheulean and Middle Stone Age technocom-

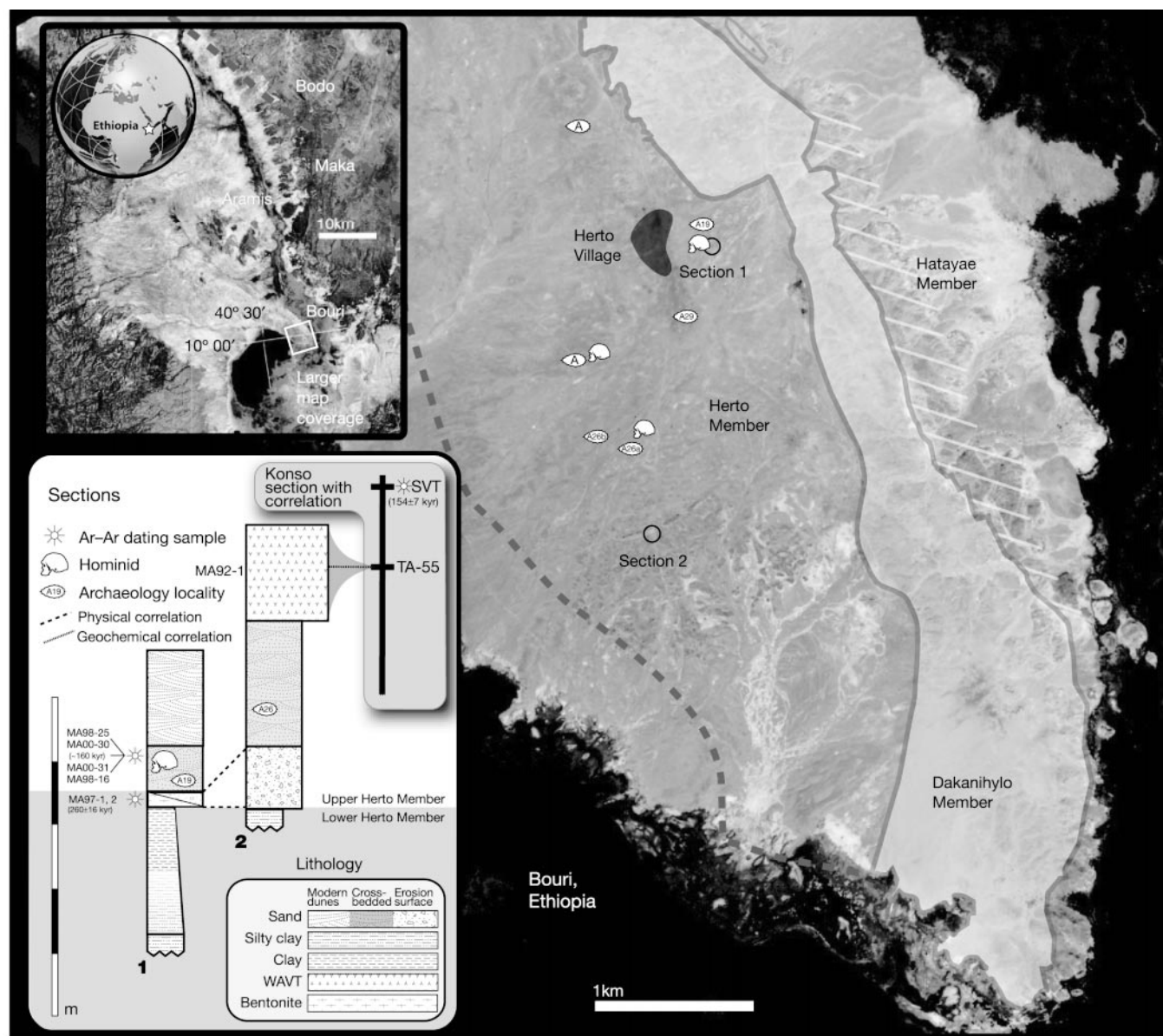
plexes. Associated faunal remains indicate repeated, systematic butchery of hippopotamus carcasses. Contemporary adult and juvenile *Homo sapiens* fossil crania manifest bone modifications indicative of deliberate mortuary practices.

Pliocene Hatayae Member sediments at the base of the Bouri Formation in Ethiopia's Afar Depression yielded *Australopithecus garhi* and the earliest evidence of the butchery of large mammals by hominids<sup>2</sup>. The overlying 1-million-year-old Dakanihylo Member deposits yielded early Acheulean assemblages associated with skeletal remains of *Homo erectus*<sup>3</sup>. The overlying Herto Member comprises two units that contain later Acheulean and early Middle Stone Age (MSA) implements<sup>4</sup>. The recovery of hominid remains in the Upper Herto Member in 1997 intensified stratigraphic, geochemical, radioisotopic, archaeological and palaeontological studies of these deposits and their contents.

The base of the Herto Member is unconformable on the underlying Dakanihylo Member sandstone. We recognize a major stratigraphic and temporal division within the 15–20-m-thick Herto Member and so designate the Upper and Lower Herto Members. The Lower Herto Member, which resulted from the uplift of the

Bouri fault block that diverted the ancestral Awash River and created Yardi Lake, consists of lignite, pinkish carbonate layers, and silty clays of predominantly lacustrine origin bearing gastropods and bivalves. The Lower Herto lithic assemblage is characterized by late Acheulean tool types; hominid remains are as yet unknown<sup>4</sup>. The age of the Lower Herto is established by an included bentonite tuff (MA97-1, 2) usually located within 1 m below the Upper/Lower Herto contact. Although contaminated by xenocrysts, it yields an <sup>40</sup>Ar/<sup>39</sup>Ar age of  $260 \pm 16$  thousand years (kyr; mean  $\pm$  s.d., here and throughout) defined by the weighted mean age of 24 (of 93) individual alkali feldspar crystals (see Supplementary Information). These Lower Herto sediments grade to fluvial and lake-margin deposits of coarse beach and yellow sandstone of the overlying Upper Herto Member, itself capped by a vitric tuff (WAVT, Waidedo Vitric Tuff, MA92-1; Fig. 1). Abundant archaeological and faunal remains are embedded in a sand unit at the base of the Upper Herto Member, the former characterized by finely made bifaces on flakes, sometimes Levallois flakes.

The contact between Lower and Upper Herto Members is defined here as a widespread erosional surface identified by a fossiliferous,



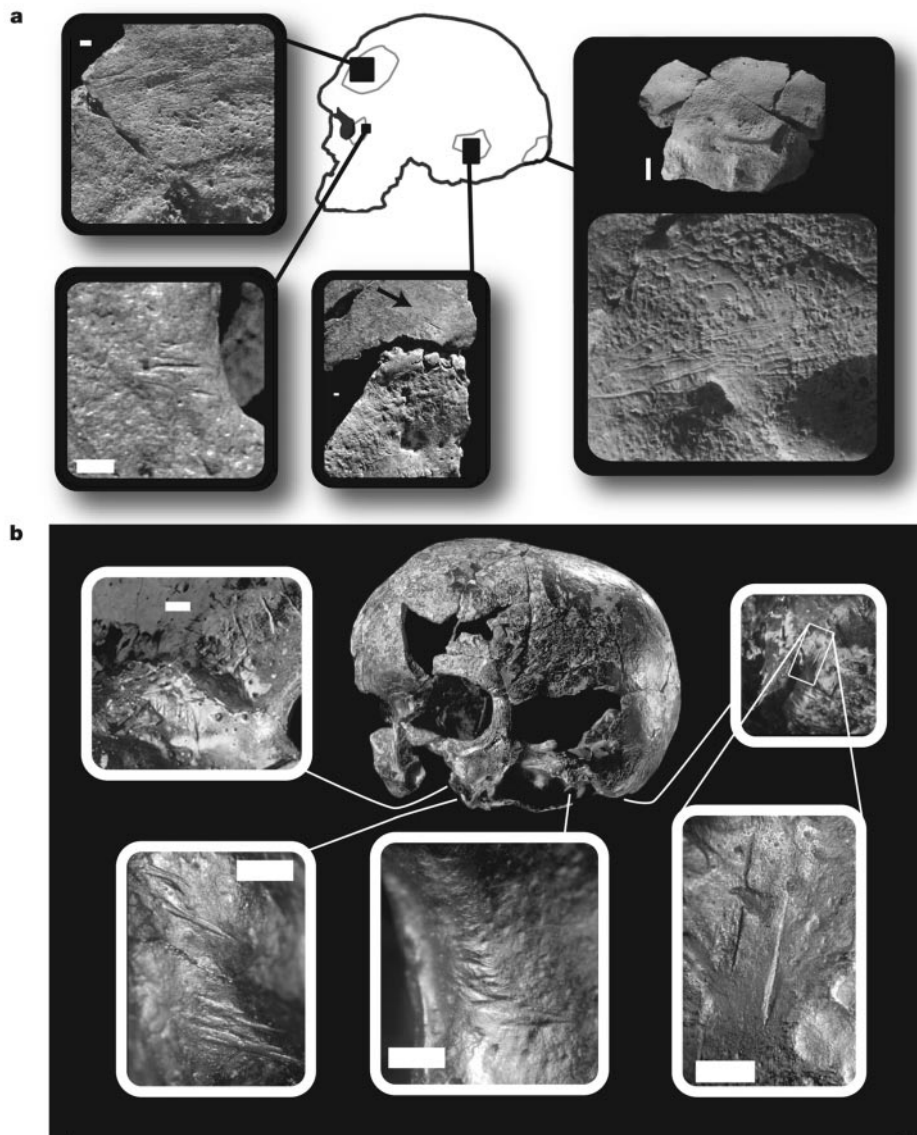
**Figure 1** Geographic and stratigraphic placement of the Herto Bouri hominid fossils and archaeological artefacts. See the text for details.

artefact-bearing, semi-indurated pebble conglomerate. This unit's consolidation renders it a widespread planar marker with distinctive geomorphological expression, recognizable by its incorporation of rounded pebbles and large bentonite clasts derived from reworking of the underlying MA97-2 volcanic unit. Small gastropod shells are often present immediately above this contact. We interpret this pebble horizon as having been deposited marginal to a shallow freshwater lake. This erosional surface is immediately overlain by a volcanoclastic sandstone and gravel deposit that yielded all of the Herto hominid fossils and all of the Upper Herto Member archaeological assemblages described below. This fluvial sand unit is variable in thickness, yellow-brown to grey, and cross-bedded, bearing abundant rolled pumices up to 15 cm in diameter. The planar pebble surface was broken in the area immediately southeast of Herto Village (Fig. 1) by syn-depositional and post-depositional faulting. Upper Herto sand deposits more than 3 m thick are preserved against the fault scarps.

Geochronological control over Upper Herto Member artefacts and fossils has been achieved through the integration of lithostrati-

graphy, tephrostratigraphy and radioisotopic dating. The maximum age for the sandstone unit containing the Upper Herto artefacts and fossils is defined by  $^{40}\text{Ar}/^{39}\text{Ar}$  ratio analyses of embedded pumices and obsidian clasts sampled from the same unit across more than 500 m (see Fig. 1 and Supplementary Information). Anorthoclase grains from two pumice clasts (MA98-25 and MA00-30) yielded plateau ages for incrementally heated multigrain samples of  $163 \pm 3$  and  $162 \pm 3$  kyr, respectively. Indistinguishable (although less precise) ages were obtained by total fusion analysis of single crystals. Multigrain incremental heating analysis of a third pumice clast (MA00-31) sampled from the same sandstone unit yielded sanidine with a plateau age of  $226 \pm 2$  kyr. Two obsidian clasts from the same sands were also analysed by incremental laser heating. One (MA98-16B) yielded a well-defined plateau age at  $160 \pm 2$  kyr.

The preponderance of pumice and obsidian with indistinguishable ages of about 160 kyr indicates that these age maxima might be close to the actual time of deposition of this fossiliferous unit. Independent confirmation of these radioisotopic determinations on



**Figure 2** Cultural modification of the Herto adult and child crania. **a**, BOU-VP-16/2 adult cranial fragments to show selected defleshing cutmarks on the left zygomatic and parietal asterionic corner, and other more superficial artificial scoring above the left temporal line

and across the occipital plane. **b**, BOU-VP-16/5 child's cranium with defleshing cutmarks on the left sphenoid, and right and left temporals, and post-mortem polish on the parietals. See the text for details. Vertical scale bar, 1 cm; horizontal scale bars, 1 mm.



the embedded volcanics was attempted by thermoluminescence dating, which proved unreliable. Uranium–thorium disequilibrium analyses of fossil bone were inconclusive.

A few hundred metres southeast and southwest of the hominid discoveries (Fig. 1), the Upper Herto fossiliferous deposits are capped by the WAVT tuff (MA92-1), at least 1.65 m thick, which is a very fine vitric ash with up to 10% finely fragmented crystals of quartz and feldspar (see Supplementary Information). The minimum age for the Herto archaeological occurrences is thus constrained by the overlying WAVT, which has resisted accurate  $^{40}\text{Ar}/^{39}\text{Ar}$  dating because of contamination by older crystals. However, major element compositions of individual glass shards determined by electron probe microanalysis, and major and trace element compositions (Supplementary Table 2) of purified bulk glass separates by direct-current argon plasma spectrometry, provide a secure correlation of the Herto Bouri WAVT (sample MA92-1) with an unnamed Pleistocene tuff (sample TA-55) from the Konso region of southern Ethiopia<sup>5</sup>. The Konso Silver Tuff (sample TG 120) found 6 m above this WAVT-correlative has been dated at  $154 \pm 7$  kyr ago, on the basis of the weighted mean of 14 individual  $^{40}\text{Ar}/^{39}\text{Ar}$  laser fusion analyses of sanidine crystals (see Supplementary Information). Eleven xenocrysts with ages more than 7 s.d. from the mean are confidently resolved from the juvenile population. On the basis of the combined stratigraphic, geochemical and radioisotopic evidence, the Upper Herto archaeological and palaeontological remains are therefore securely constrained to be between  $160 \pm 2$  and  $154 \pm 7$  kyr old. They are thus late Middle Pleistocene in age, a placement entirely consistent with the palaeontological and archaeological evidence.

Chronometric placement of the Herto discoveries is significant because the accurate dating of faunas and artefacts of many sites of this general antiquity in Pleistocene Africa has proved notoriously difficult<sup>6</sup>. Consequently, the transition to anatomical modernity in African hominid populations has proved difficult to document because of the lack of chronological control over the widely scattered hominid fossils that are often singular and fragmentary<sup>7</sup>. Compounding this problem is the developing consensus that the transition between the Acheulean and MSA technocomplexes in Africa was both temporally and spatially complex, involving blades, points and Levallois elements in late Acheulean assemblages, and including bifaces persisting in assemblages that would otherwise be termed MSA<sup>8,9</sup>. The antiquity and geographic position of the Herto discoveries in the Horn of Africa therefore advance our understanding of both technological and anatomical change.

Upper Herto Member archaeological assemblages have been observed across more than 5 km of modern landscape. Artefacts and stratigraphically associated fossils are particularly abundant in the fluvial deposits immediately atop the widespread erosional surface defining the Lower to Upper Herto submember boundary as described above. The bulk of the vertebrate fauna also comes from this widespread sand unit and includes a derived extinct bovine, *Kobus* species, *Thryonomys*, *Hippopotamus*, *Equus* and *Connochaetes*. These taxa indicate the proximity of both aquatic and grassland habitats.

The Upper Herto archaeological assemblages described here were recovered mainly from controlled surface collections of representative lithic artefacts at the following localities (all handaxes, flake tools, cores, flakes and blades were collected on sublocalities designated at each archaeological occurrence; see Fig. 1 and Supplementary Table 3): BOU-A19 (71 artefacts from sublocalities A, C–G, H–N; total 10,100 m<sup>2</sup>); BOU-A26 (331 artefacts from sublocalities A–C; total 777 m<sup>2</sup>); and BOU-A29 (194 artefacts from 5,000 m<sup>2</sup>). All of these surface assemblages were judged to derive from the lower 1 m of the hominid-bearing sand unit described above and were recovered from erosional exposure of *in situ* artefacts, or from controlled surface collections made on the basis of preservational characters, adhering matrix and/or geomorpho-

logical disposition. In addition, to control these surface collections further with additional *in situ* materials, we undertook excavations at sublocality BOU-A19B, the site of discovery of a cutmarked hippopotamus cranium. From this excavation we recovered 29 lithic artefacts interbedded with a rich faunal assemblage from less than 20 cm of archaeological deposit excavated across 33 m<sup>2</sup>. Another excavation at the location (BOU-A19H) of the *in situ* discovery of the complete adult hominid cranium described below revealed 15 artefacts from less than 20 cm of archaeological deposit with abundant modified faunal remains, excavated across 41.5 m<sup>2</sup>. Excavated remains differed in no significant manner from the controlled surface collections, and there was no evidence of any contamination from younger deposits (only thin aeolian dunes now cover the sampled unit in the area).

The combined Upper Herto archaeological assemblages vary spatially in their lithological and typological contents. The Levallois method is well represented across samples. Levallois and smallish Levallois flakes and points normally associated with the African MSA are present, as are Acheulean cleavers and other bifaces. All these tool types are represented by examples found *in situ* in the hominid-bearing sand unit. Similar assemblages are traditionally classified as final or 'transitional' Acheulean.

Controlled surface collection and excavation at Upper Herto localities BOU-A19, BOU-A26 and BOU-A29 yielded a pooled lithic assemblage of 640 analysed artefacts (see Supplementary Information) adequate to characterize the industry that is stratigraphically contemporary with the hominid fossils. Whereas the pooled assemblage assessed here is adequate to characterize the Upper Herto Member archaeology qualitatively, additional excavations will be required for a fuller quantitative analysis of the lithic materials. Even at this early stage of investigation it is evident that handaxes and picks are rare (less than 5% of flaked implements collected), as are blades (less than 1%). The rarity of handaxe preparation flakes among the collected lithics indicates that these tools may have been made elsewhere. Raw material is predominantly fine-grained basalt, except for points and blades, which were made mostly on obsidian. Cryptocrystalline rocks used for some scrapers are rarer, comprising 5% of the total retouched lithics.

The Levallois method is well represented in the pooled assemblage and was used frequently in the production of the handaxes and cleavers. Evidence of Levallois method is also observed on 48 flakes, blades and points. Of the 63 tools on flakes, 9 were made on Levallois flakes. Preparation was usually radial centripetal. Flakes are mostly elliptical with flat section, with platforms almost always faceted convex. They are typologically 'chapeaux de gendarme' and 'en aile d'oiseau', with a platform angle always between 90° and 95°. Of the 53 cores collected and analysed, 28 are discoid; these are probably the result of exhaustive exploitation of the Levallois cores. Blade technology is present, but rare, with only four blades (three *in situ*), all on obsidian.

The analysed 28 bifaces span a wide size range and were all made on fine-grained basalt. They are represented by ovates, elongate ovates, triangulars, cleavers, and a pick, biface scraper and biface nucleus. The 17 handaxes with ovate and elongate ovate plan forms were always made on flakes and finished with soft hammer technique. Edges are regular and show secondary edge retouching.

Of the 25 side scrapers, simple side scrapers dominate ( $n = 22$ ), the remainder being convergent side scrapers and a double-sided scraper with biconvex edges. Most specimens show direct retouch. Fifteen end-scrapers or rabots are present, made on thick flakes that sometimes conserve cortex. The working edges are usually convex and regular, made with lamellar retouching. Most of these tools are also side scrapers; some are carinated and resemble Aurignacian types.

The technological and typological positions of the Upper Herto assemblages seem closest to Garba III at Melka Kunture, Ethiopia<sup>10,11</sup>. As at Herto, Garba III includes terminal Acheulean han-

daxes, typical Levalloisian method, and many retouched tools on flakes (side-scrapers and end-scrapers, backed knives, burins, unifacial and bifacial points). The Garba III assemblage has been considered transitional between the Acheulean and the MSA<sup>11</sup>.

Indications at Gaddemota (Ethiopia) and Kapthurin (Kenya) are that by about 280 kyr ago, Acheulean assemblages at some African sites had incorporated new elements typical of the MSA. The demonstrably younger Herto Member assemblages, with their mix of Acheulean and MSA technological attributes and artefact forms, add complexity to this picture and confirm that the transition to the MSA in eastern Africa was not a simple or a gradual process<sup>8,12</sup>. The cultural mosaic across time and space probably derives from variations in geography, ecology and raw material, and other cultural variables that will require disentanglement by further research.

The Upper Herto hominids occupied the margin of a freshwater lake, and archaeological evidence from the base of this submember indicates hominid butchery of large mammal carcasses, particularly hippopotamids. Such butchery is documented for the earlier Lower Herto beds<sup>4</sup>. It persisted in the Upper Herto, where numerous examples of cutmarked and percussion-fractured hippopotamus and bovine remains are observable in the sand unit directly above the erosional surface. One occurrence shows abundant remains of several hippo calves, mostly newborn to a few weeks old, scattered together with butchered adults.

The basal sandstone of the Upper Herto Member yielded three hominid crania found just east of Herto village. The most intact is BOU-VP-16/1, an adult cranium with dentition found *in situ* by D. DeGusta on 27 November 1997. Its left side was partly lost owing to modern erosion, whereas its intact right side faced downwards and was preserved. The cranium was cemented in place by indurated sand lenses. Excavation demonstrated conclusively that the specimen was not intrusive. A rich archaeological horizon was encountered 25 cm below the cranium, but there was no direct archaeological association within the 60 m<sup>2</sup> excavation centred on the fossil. The BOU-VP-16/5 juvenile individual was a surface find made on 3 December 1997 by B. Asfaw. Adhering sand matrix shows that it was also embedded in stained sands immediately above the erosional surface. The third major hominid fossil from contemporary Upper Herto sediments was another adult (BOU-VP-16/2), this individual represented by 24 mostly vault fragments found on the surface of the erosional contact by C. Pehlevan on 27 November 1997. No hominid postcranial elements have yet been recovered from Upper Herto Member deposits.

The anatomical and evolutionary significance of the hominid crania is considered elsewhere<sup>1</sup>. The three primary Herto crania all bear cultural modification indicative of mortuary practice. The least modified is the more intact male individual BOU-VP-16/1, which shows a long, thin, weakly sinuous, 35-mm vertical cutmark on the anteroinferior corner of its right parietal, and a second, shorter mark surmounting the supramastoid crest of the right temporal. No other modification is visible. The second adult cranium, also a probable male, was highly fragmented and scattered after it emerged from the same sand unit. The lack of recovered dental, facial or basicranial parts indicates that it may have been embedded as a calotte. Evidence of intensive bone modification is present on 15 of its 24 recovered fragments. Some exhibit cutmarks that are probably associated with removal of soft tissue. Deep, typical defleshing cutmarks are seen on the parietals, left zygomatic, frontal and occipital (Fig. 2). More abundant but more superficial marks showing a repetitive scraping motion are present around the vault circumference, above the nuchal and temporal lines. The latter pattern of bone surface modification is almost never present in hominid or nonhuman faunal remains processed for consumption<sup>13</sup>, and is therefore unlikely to represent evidence of utilitarian or economic behaviour.

The juvenile cranium lacks such superficial scoring marks but

displays an unambiguous series of defleshing-related cutmarks on the basicranial surfaces of its sphenoid and temporal bones on both right and left sides. These fine but deep and repetitive cutmarks (Fig. 2) were made by a very sharp, probably obsidian flake edge. Their locations, dimensions and directions around the perimeter of the glenoid fossae indicate that this defleshing manipulation must have occurred after removal of the mandible. The intentional and deliberate removal of soft tissues such as basicranial vessels, nerves and muscles is therefore indicated. The specimen lacks the entire occipital region surrounding the foramen magnum, and the edges of this broken region are smooth and polished, as are the specimen's unweathered parietal surfaces.

Ethnographic and osteological evidence from several cultures documents the post-mortem manipulation and curation of human remains as part of mortuary practices<sup>13</sup>. For example, some New Guinean crania show cutmarks, decoration and polishing reminiscent of traces seen on the Herto hominids<sup>14</sup>. The earliest indications of cultural modification of hominid remains are the disarticulation-related cutmarks present on the early *Homo* specimen Stw-53 from Sterkfontein<sup>15</sup>. The Middle Pleistocene Bodo cranium, which is much earlier than the Herto hominids, gives the earliest evidence of non-utilitarian mortuary practice in the hominid fossil record, but the cutmarks it bears seem to be related to defleshing and not decoration<sup>16</sup>. The diverse bone modifications marking the three most intact Middle Awash Herto hominid crania indicate post-mortem defleshing with stone tools in the Upper Pleistocene, in an archaeological context straddling the Acheulean and MSA. Polishing and intentional scraping modifications evident on two of these crania indicate that the Upper Herto hominids may have manipulated the crania of their dead in mortuary practices whose dimensions, context and meaning might be revealed only by further discoveries. □

Received 21 November 2002; accepted 14 April 2003; doi:10.1038/nature01670.

- White, T. D. *et al.* Pleistocene *Homo sapiens* from Middle Awash, Ethiopia. *Nature* **423**, 742–747 (2003).
- Asfaw, B. *et al.* *Australopithecus garhi*: A new species of early hominid from Ethiopia. *Science* **284**, 629–635 (1999).
- Asfaw, B. *et al.* Remains of *Homo erectus* from Bouri, Middle Awash, Ethiopia. *Nature* **416**, 317–320 (2002).
- DeHeinzelin, J., Clark, J. D., Schick, K. D. & Gilbert, W. H. (eds) *The Acheulean and the Plio-Pleistocene Deposits of the Middle Awash Valley, Ethiopia* (Royal Museum of Central Africa, Tervuren, Belgium; Ann. Sci. Geol. **104**, 2000).
- Katoh, S., Nagaoka, S., WoldeGabriel, G., Beyene, Y. & Suwa, G. Preliminary study on geomorphological development since the early Pleistocene in Konso-Gardula area (KGA), southern Ethiopia. *Proc. Gen. Meet. Assoc. Jpn. Geogr.* **49**, 208–209 (1996).
- Howell, F. C. Paleo-demes, species clades, and extinctions in the Pleistocene hominin record. *J. Anthropol. Res.* **55**, 191–243 (1999).
- Stringer, C. B. in *The Age of the Earth: From 4004 BC to AD 2002* (eds Lewis, C. L. E. & Kneel, S. J.) 265–274 (Geological Society of London, Spec. Publ. **190**, 2001).
- Tryon, C. A. & McBrearty, S. Tephrostratigraphy and the Acheulean to Middle Stone Age transition in the Kapthurin Formation, Kenya. *J. Hum. Evol.* **42**, 211–235 (2002).
- Clark, J. D. in *A Very Remote Period Indeed: Papers on the Palaeolithic Presented to Derek Roe* (eds Milliken, S. & Cook, J.) 1–18 (Oxbow Books, Oxford, 2001).
- Hours, F. in *Proceedings of the Seventh Pan-African Congress of Prehistory and Quaternary Studies, 1971, Addis Ababa* (eds Abebe, B., Chavaillon, J. & Sutton, J. E. G.) 99–104 (Ministry of Culture, Addis Ababa, 1976).
- Chavaillon, J., Chavaillon, N., Hours, F. & Piperno, M. From the Oldowan to the Middle Stone Age at Melka-Kunturé (Ethiopia): Understanding cultural changes. *Quaternaria* **21**, 87–114 (1979).
- Clark, J. D. in *Hominid Evolution: Lifestyles and Survival Strategies* (ed. Ullrich, H.) 277–292 (Edition Archaea, Berlin, 1999).
- White, T. D. *Prehistoric Cannibalism at Mancos 5MTUMR-2346* (Princeton Univ. Press, 1992).
- White, T. D. & Toth, N. The question of ritual cannibalism at Grotta Guattari. *Curr. Anthropol.* **32**, 118–138 (1991).
- Pickering, T., White, T. D. & Toth, N. Cutmarks on a Plio-Pleistocene hominid from Sterkfontein, South Africa. *Am. J. Phys. Anthropol.* **111**, 579–584 (2000).
- White, T. D. Cutmarks on the Bodo cranium: A case of prehistoric defleshing. *Am. J. Phys. Anthropol.* **69**, 503–550 (2000).

Supplementary Information accompanies the paper on [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** J.D.C., a legend in African archaeology, founded and co-led the Middle Awash project until his death in February 2002. We thank A. Almquist, A. Asfaw, M. Asnake, T. Assebework, D. Brill, J. DeHeinzelin, A. Getty, Y. Haile-Selassie, A.-R. Jaouni, B. Latimer, C. Pehlevan, K. Schick, S. Simpson, P. Snow and Y. Zeleka for fieldwork and analytical studies; the Earth Environmental Sciences Division, Los Alamos National Laboratory, for access to electron microprobe and other support; D. DeGusta, F. C. Howell, C. O. Lovejoy, L. Hlusko, F. Bibi, R. Klein, L. Jellema and E. Vrba for review and/or assistance; and J. Feathers, and J. Westgate and A. Sandhu for assessing the feasibility of luminescence and fission track dating, respectively, on some of the tephra. We thank the Ministry of Youth, Sports and Culture, the Authority for Research and Conservation of the Cultural Heritage, and the National Museum of Ethiopia for permissions; the Afar Regional Government and the Afar people of the Middle Awash, particularly the Bouri-Modaitu community and H. Elema; and many additional individuals for contributions. This research was supported by the NSF (US), the Institute of Geophysics and Planetary Physics (University of California at Los Alamos National Laboratory), and the Japan Society for the Promotion of Science. Additional financial contributions were made by the Hampton Fund for International Initiatives, Miami University.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to T.W. (timwhite@socrates.berkeley.edu).

## The remarkable inefficiency of word recognition

Denis G. Pelli\*, Bart Farell† and Deborah C. Moore†

\* Psychology and Neural Science, New York University, New York, New York 10003, USA

† Institute for Sensory Research, Syracuse University, Syracuse, New York 13244-5290, USA

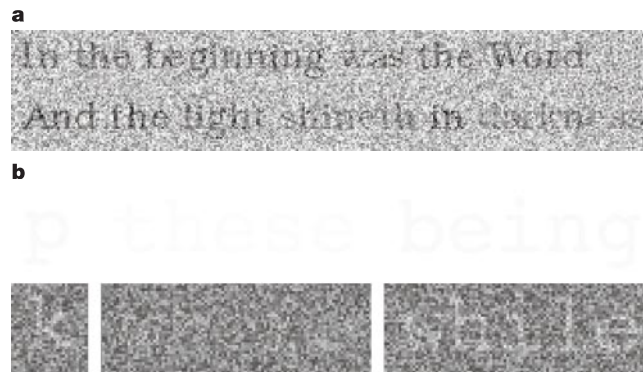
Do we recognize common objects by parts, or as wholes? Holistic recognition would be efficient, yet people detect a grating of light and dark stripes by parts. Thus efficiency falls as the number of stripes increases, in inverse proportion, as explained by probability summation among independent feature detectors<sup>1</sup>. It is inefficient to detect correlated components independently. But gratings are uncommon artificial stimuli that may fail to tap the full power of visual object recognition. Familiar objects become special as people become expert at judging them<sup>2,3</sup>, possibly because the processing becomes more holistic. Letters and words were designed to be easily recognized, and, through a lifetime of reading, our visual system presumably has adapted to do this as well as it possibly can. Here we show that in identifying familiar English words, even the five most common three-letter words, observers have the handicap predicted by recognition by parts: a word is unreadable unless its letters are separately identifiable. Efficiency is inversely proportional to word length, independent of how many possible words (5, 26 or thousands) the test word is drawn from. Human performance never exceeds that attainable by strictly letter- or feature-based models. Thus, everything seen is a pattern of features. Despite our virtuosity at recognizing patterns and our expertise from reading a billion letters, we never learn to see a word as a feature; our efficiency is limited by the bottleneck of having to rigorously and independently detect simple features.

The role of components in object recognition is still mysterious<sup>4-7</sup>. For most objects we don't even know what the components are, but we do know that words are made of letters. Is a familiar word recognized as an image, or as a combination of individually recognized letters? Despite a century of careful study<sup>8-16</sup>, it has never been noted that these two alternatives predict very different thresholds for identifying words. Once we have defined a few terms, you can test the predictions with your own eyes.

The strength of visual signals is traditionally specified by 'contrast', which here is the ratio of the luminance increment of the letter or word to the background luminance. However, for ideal-observer analysis it is helpful to specify 'contrast energy', which for a letter or a word is the product of squared contrast and 'ink' area. In general, contrast energy is the integral of the squared signal contrast over the extent of the stimulus. Energy matters: mathematical work on radar in the 1950s proved that the energy of a known signal completely determines its detectability in white noise<sup>17</sup>. 'Threshold', the border between seeing and not seeing, is defined here as the contrast or energy required by the observer to correctly identify the letter or word 64% of the time.

In Fig. 1a, the two lines of faint text have the same overall contrast energy, but differ in the way that the energy is distributed. To optimize recognition by parts, the first line gives each letter the same energy. To optimize recognition as wholes, the second line gives each word the same energy. Recognition by parts predicts that the longer words in the second line will be illegible, as you see, because there isn't enough energy per letter. Recognition as wholes predicts that all words in the second line will be equally legible, contrary to what you see. Figure 1b shows the predictions for your word threshold on blank and noisy backgrounds. The same analysis applies to both backgrounds because the observer effectively adds his or her intrinsic visual noise to the display<sup>18</sup>.

Our recognition-as-wholes predictions are based on the ideal observer, which, given the stimulus and its statistics, achieves the best possible expected performance by choosing the most probable letter or word<sup>17,18</sup>. We can assess human performance on an absolute scale by defining 'efficiency' as the ratio of the ideal's threshold energy to the human observer's: the fraction of the energy used by



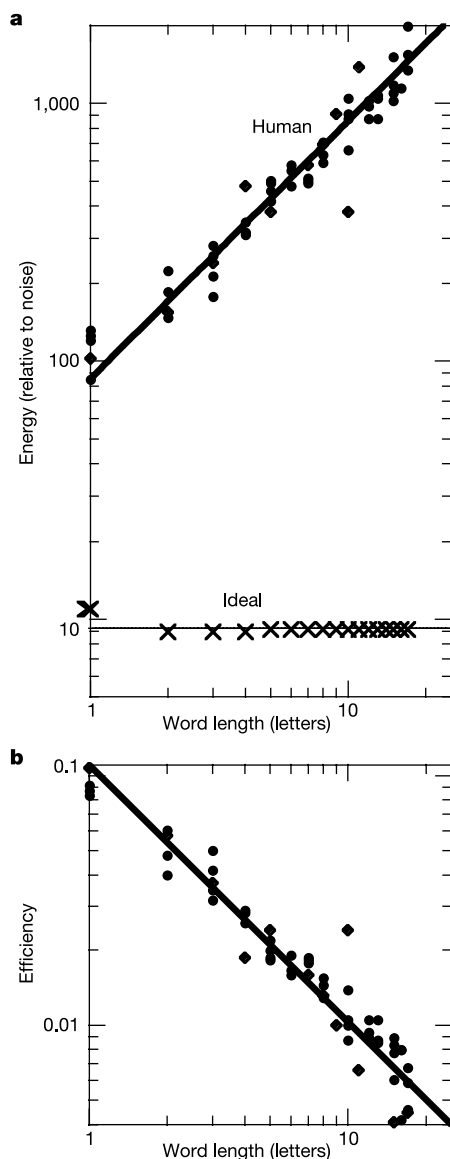
**Figure 1** By letter or by word? **a**, Both lines of the quotation have the same total contrast energy. In the first, the energy is divided equally among the letters. In the second, the energy is divided equally among the words, regardless of length. In principle, at a given noise level a pattern's detectability depends only on its energy, but in the second quotation the short words pop out and the long words disappear. This word-length effect shows that human readers cannot efficiently integrate the energy across a whole word. (The quotation reads 'In the beginning was the Word ... And the light shineth in darkness'.) **b**, Two predictions for the word threshold. The left column shows a letter at threshold on a uniform white background (top row) and on a noisy background (bottom row). They may take a minute to appear. The middle column shows a 5-letter word at the same energy ( $1/\sqrt{5}$  the contrast), which is the threshold predicted by recognition as wholes, and the right column shows a word at 5 times the energy (the same contrast), which is the threshold predicted by recognition by parts. The words in the middle column would be identifiable if you could see words as efficiently as you see letters. (The first line reads 'p', 'these', 'being', and the second reads 'k', '?????', 'while', where the identity of the middle word remains undisclosed.) Note: The faint lettering on a white background is at the limits of what can be rendered on the printed page. The PDF of this letter prints the figure successfully on most modern printers, especially colour printers. If in doubt, readers are urged to refer to a more robust version of Fig. 1b available as Supplementary Information, which accommodates variations in printers' rendering and readers' sensitivity.



the observer that is ideally required to account for the observer's measured performance<sup>18,19</sup>. This strips away the intrinsic difficulty of the task, exposing a pure measure of human ability.

Our argument begins with our finding that human efficiency for word identification is inversely proportional to word length, independent of the number of possible words, as predicted by recognition by parts with suppression of weak signals, 'squenching'. Then we show that, notwithstanding the well known 'word superiority effect', human word identification never exceeds the accuracy attainable by strictly letter-based models. Related work indicates that letters are made of features. Finally, we conclude that visual recognition of even the most familiar objects is severely restricted by a first stage of independent feature detectors that squelch and each integrate no more than a letter, and probably much less.

The word-length effect, for words 2–16 letters long, is shown in Fig. 2 for two observers. Figure 2a shows that their threshold energy is proportional to word length, whereas the ideal observer's



**Figure 2** The effect of word length. **a**, Threshold energy for identifying one of 26 words, as a function of word length. The data at length 1 are for single letters. The ideal observer's thresholds (crosses) lie near the zero-slope line. The human observers' thresholds (JP, circles; JG, diamonds) lie near the unit-slope line. **b**, Efficiency (the ratio of ideal and human thresholds) derived from **a** as a function of word length. The points are close to the line with  $-1$  slope: efficiency is inversely proportional to word length.

threshold is practically independent of word length. Figure 2b shows that efficiency, the ratio of ideal and human thresholds, is inversely proportional to word length. Efficiency for  $n$ -letter words is  $1/n$  that for single letters. This result is not surprising for longer words, as observers take in only a modest number of letters in a glimpse, perhaps 4.5 (ref. 20), which predicts the reciprocal drop in efficiency for longer words, as the ideal observer uses all the letters. What is surprising and important about Fig. 2 is that the same slope extends left to the shortest words, and even single letters. Thus, the required energy per letter is independent of word length.

This is not merely a consequence of just size or contrast. Although a word is bigger than a letter, and words and letters have the same threshold contrast, the human limitation demonstrated in Figs 1 and 2 is neither an inefficiency for all large objects nor a fixed threshold contrast that all objects must exceed to be seen. Increasing letter size fivefold, to match the width of a word, reduces the letter's threshold contrast, in noise, fourfold<sup>21</sup>. Because seeing the large letter requires only a fraction of the contrast required to see a similar-width word, any explanation of the poor visibility of words must penetrate past their size to consider their internal structure. Our results indicate that, rather than directly recognizing complex familiar objects, such as words, our visual system detects smaller components—letters or perhaps features of letters—and only then recognizes the object specified by these components. Nothing is seen unless its components are detected<sup>22,23</sup>.

One might ask whether the human observer could reasonably be expected to confine his or her word search to the 26 words used in the test, as the ideal does. Our design minimized this concern by using the 26 most common  $n$ -letter words, and displaying them as the response alternatives in every trial. Identifying letters independently (recognition by parts) is inefficient because the letters in a word are correlated. That's because the number of  $n$ -letter words, 26 in our experiment or thousands in real life, is a tiny fraction of the number of possible strings,  $26^n$ . Still, one might worry that this human inefficiency is an artefact of using only 26 words. We addressed these concerns by measuring human and ideal thresholds for identification of one of 2,213 five-letter words (the most common 2,213), presenting each word at the same relative frequency as it appears in print<sup>24</sup>, and found the same efficiency (approximately 4%) as for the 26 most common. (These efficiencies, for experienced observers with Courier font, are slightly higher than in Fig. 2, which is for new observers with Bookman font. The 100,000-trial experience, in various conditions, and the Courier font both contribute to the  $\times 1.5$  higher efficiency<sup>21</sup>.) Finding the same efficiency means that human and ideal thresholds are affected equally by word frequency. Thinking that perhaps only extremely common words enjoy recognition as elementary visual patterns<sup>25</sup>, we measured human and ideal thresholds for identifying the five most common 3-letter words (the, and, was, for, his). The least frequent of these words (his) is encountered 100 times in an hour of reading, nearly 400,000 times in a decade of reading an hour a day. Even so, the five most common 3-letter words yield practically the same efficiency as the 26 most common (4.8% compared to 4.5%) and reciprocity holds: the efficiency for identifying the five most common 3-letter words is  $1/3$  that for single letters: 4.8%/15%. Changes in intrinsic task difficulty—5, 26 or 2,213 alternatives—invalidate comparison of raw thresholds unless modelling assumptions are made, but efficiencies are always directly comparable.

The ideal observer chooses the best-matching template, using one template for each possible word. The root mean square (r.m.s.) difference between the stimulus and the template is a measure of the likelihood that the stimulus is the template plus noise. To choose the most probable word the ideal observer weighs each word's likelihood by its frequency. Template matching integrates energy efficiently over the extent of the template. If humans did template matching, with accurate templates for words of every length, then the slope in Fig. 2b would be zero, not  $-1$ . Figure 2 indicates the

absence of templates of more than one letter, as the energy beyond one letter doesn't reduce the required energy for the first letter.

The brain is well equipped to do template matching. A typical neuron sums over 10,000 synapses, each with a different gain. Any neuron that integrates over part of the visual field computes the likelihood of the presence of a signal matching the neuron's sensitivity profile. Thus, neurons with very simple receptive fields have been called "fly detectors"<sup>26</sup>. One can speculate that there might be neurons (perhaps in brain area IT) that linearly integrate over more complex receptive fields that match a face, letter or word, but such neurons would allow the observer to attain much higher efficiencies than found here, placing their existence in doubt.

In principle, identifying a letter independently requires the same energy for that letter as would be required to identify the whole word (if there are, as in Fig. 2, the same number of possible letters and words). The fact that a word is unreadable by our observers unless its letters are separately identifiable is evidence for recognition by components; that is, identification of the word is mediated by independent detection of components that are a letter or less. We define 'features' as image components that are detected independently, unaffected by the presence of other features. Independent feature detection is a bottleneck, especially when feature thresholds are high; complex objects will be visible only when the energy per feature reaches threshold.

Our data indicate that there are no multi-letter features. Efficiency for letters is independent of age after ten, only weakly dependent on size and alphabet (English, Armenian, Devanagari and Hebrew), and inversely proportional to complexity<sup>21</sup>. 'Complexity' is a scale-invariant physical measure: perimeter squared over 'ink' area<sup>21</sup>. The number of features in a letter may be proportional to its complexity. It seems that in identifying letters, observers use no feature more complex than the average letter in the simplest alphabet tested<sup>21</sup>, about one-third the complexity of the Bookman and Courier fonts used here. As letters and words are designed to be legible, and the reader's visual system presumably has adapted itself to them as much as it can, we conclude that the feature bottleneck is unavoidable. Objects are recognized by means of independent detection of their component features, which are much simpler (less complex) than a single Bookman letter.

The efficiency result—the reciprocal relation between efficiency and number of components—is secure and assumption-free, but how general is it? We extended it from obscure gratings<sup>1</sup> to common words. It also applies to letters if we suppose that complexity is proportional to the number of features. But these stimuli were all at threshold. Threshold stimuli are directly relevant to real-life reading of highway signs, which are usually read at great distance as soon as they become readable. Might ordinary reading, at high contrast, be mediated by different mechanisms? That seems unlikely. Critical-band masking studies have characterized the channels (feature detectors) that mediate letter identification at threshold<sup>27</sup>. Experiments at high supra-threshold contrasts, measuring the effect of noise on reading rate, reveal the same channel tuning<sup>28</sup>.

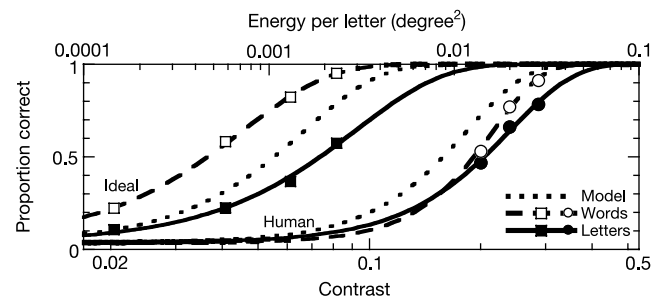
We usually see things quite reliably or not at all, with a fairly abrupt transition between the two. Indeed, the 'psychometric function', the probability of seeing, rises much more steeply as a function of contrast than predicted by theory of signal detectability for an exactly known signal<sup>17</sup>. The steep psychometric function means that weak signals are suppressed. Engineers call this 'squenching': in better walkie talkies, a nonlinear analog circuit turns down the volume when the signal is weak relative to the background noise, to cut out the hiss when no one is speaking. In the same way, human vision squelches features, allowing them to pass only if they are well above the noise. We call this 'detecting rigorously'. It achieves reliability at the expense of efficiency. The human visual system has a vast number of feature detectors, each of which can raise a false alarm, mistaking noise for signal. Squelching blocks the intrusion of countless false features that would besiege us if weak features were

not suppressed. However, it impairs our ability to see more complex objects, like words. In the limiting case, squelchers pass all signals above a certain threshold contrast, and none below. Whether the squelch is gradual or abrupt, it is, in effect, deciding whether or not a signal is present and acting on that decision.

Discussions of letter and word recognition typically suppose various stages in the recognition process: identification of features, then letters, then the word<sup>14,29</sup>. One approach to modelling how an observer performs a task is to specify just the first stage of processing, leaving subsequent stages unspecified. However, such models must avoid the common mistake of specifying a transformation that preserves all task-relevant information. Such a transformation can always be undone by the subsequent unspecified stages, so it does not constrain performance, making the model psychophysically inconsequential and untestable. Proposing a first stage constrains performance only to the extent that the stage discards task-relevant information. In that spirit, when we assert that a part of a model, or the brain, 'makes a decision', we mean not only that it passes that decision on, but also that it discards (fails to pass on) the information the decision is based on. Thus, we consider the idea that all word and letter identifications are strictly letter-based; that is, they depend on the visual stimulus solely through a stage that identifies each letter independently and passes on only that identification, discarding the rest of the stimulus information. We leave the later stages, after letter identification, unspecified. Each decision is a bottleneck.

Selfridge's Pandemonium Model<sup>29</sup> cascades row after row of 'demons' that compute the likelihood of features, then letters, and then words, achieving perfect efficiency by not discarding any choice-relevant information until the final shouting match. It is an ideal observer. Similarly, the Interactive Activation Model<sup>14</sup> computes each possible letter's likelihood at each letter position, and postpones the information-discarding shouting match until the final response-selection stage. Presumably, given optimal weights and extended to report whole words, it would be nearly ideal, so it too would have similar energy thresholds for words and letters, unlike what we report here for human observers.

It is a simple matter to calculate how accurately a word can be identified by a strictly letter-based observer. The observer's first stage identifies the letters independently. These identifications are tentative. To perform optimally, the observer, having the string of tentatively identified letters, uses a historical table of his or her single-letter confusion probabilities (identifying one letter as



**Figure 3** Proportion correct in identifying a letter (filled symbols) or a 5-letter word (open symbols) in noise as a function of contrast. Average energy per letter is indicated on the upper x axis. This is for one human, WT (3,000 trials per point), and the ideal observer (100,000 trials per point). Note that the human's contrast thresholds (64% correct) for letters and words are similar, whereas the ideal observer's are very different. For each observer, the dotted line is the best possible strictly letter-based word identification performance, given the observer's measured single-letter confusion probabilities (of identifying each letter as another) at that contrast. (A second human observer, DM, gave very similar results, not shown.) All curves are maximum-likelihood Weibull fits ( $\gamma = 1/26$ )<sup>21</sup>.

another) at that contrast to choose the most probable word from the list of alternatives. (The ideal, described earlier, computes each possible word's likelihood based on the stimulus. Our letter-based two-stage observer is not ideal, and must settle for likelihood based on the independent letter identifications, rather than the stimulus itself.)

The performance of this best-possible letter-based observer is plotted as the dotted curves in Fig. 3. The left curve is based on the ideal's single-letter confusion probabilities at each contrast, and the right curve is based on the human's. Each dotted curve is the best possible accuracy for strictly letter-based word identification by that observer, given the observer's measured letter performance. Note that the human's word performance (right dashed line) never exceeds this letter-based bound, whereas the ideal's word performance (left dashed line) far exceeds it. Thus, despite reading for decades, a hundred million words<sup>21</sup>, our observers identify even the most common words with an accuracy attainable through strictly letter-based identification.

Humans squelch features. The ideal observer does not. Nor does the two-stage model based on the ideal letter identifier. It is inefficient to detect correlated components independently, and this is exacerbated by squelching. We can calculate efficiency from the threshold contrasts of the curves in Fig. 3 for lengths 1 (letter) and 5 (word). Efficiency of the model, which doesn't squelch, drops as words grow longer. Efficiency of the human, who does squelch, drops more, as the reciprocal of word length.

At first sight, our conclusion may seem incompatible with the hundred-year-old 'word superiority effect', whereby a letter within a word is recognized better than a letter presented in isolation or in a scrambled word<sup>8–16</sup>. The word context improves letter identification even when it provides no information that can distinguish between the response alternatives. For example, *coin* versus *join* is easier than *c* versus *j*. Thus, in Fig. 4, the proportion correct for words (dashed line) is higher than that for letters (solid line). But is the observed context effect big enough to prove that the internal letter identifications are not independent of each other? No, it's too small. The human performance plotted in Fig. 4 is consistent with strictly letter-based word identification. The dotted line shows how well the observer would perform by first identifying the test letter as a ... z, and then choosing the more likely of the two response alternatives (for example, 'c' or 'j') based on her own single-letter confusion probabilities. This is the best possible strategy for an observer who is strictly letter-based. This letter-based upper bound applies to both letter and word conditions, and in fact is above both. The worse human performance shows that the observer is not using the optimal strategy, quite possibly because she doesn't know her own

single-letter confusion probabilities and thus fails to choose the most probable letter or word (c or j) given her tentative internal letter identification (a ... z). The strictly letter-based upper bound is also above the word advantage found in other studies, which find the proportion correct for letter identification to be 0.05–0.15 greater in a word context<sup>10–16</sup>. The word context (dashed line) improves performance, bringing it closer although still not exceeding the (dotted) upper bound. The slightly higher human performance in a word context (the word superiority effect) suggests that observers more accurately incorporate their historical confusion probabilities when reading words than when identifying letters, which is consistent with the observation that observers perform better if they attend to the entire word rather than just to the target letter<sup>12</sup>. Perhaps years of fast reading have trained the second-stage word-recognition process to learn the observer's letter-confusion probabilities, to more efficiently map strings of tentatively identified letters to real words. Figure 4 shows that the 0.10 upward increase in accuracy corresponds to a factor of  $\times 1.15$  leftward reduction in threshold contrast ( $\times 1.3$  in energy). Thus, there are not one but two effects, one small and one large. The word superiority effect increases efficiency by a mere factor of  $\times 1.3$ . The word-length effect, described here, is big, reducing efficiency by the word length,  $\div 5$  for a 5-letter word. Both effects are consistent with strictly letter-based word identification.

In evaluating whether recognition is by parts or as wholes, we took a 'best of breed' approach, considering the best possible performance of models that conform to simple assumptions about the observer's internal processes. Our analysis has focused on letters because they are obvious components of words. More generally, the same approach may be applied to critically test any conjecture that object recognition is mediated by independent decisions about a specified set of features or components. The components in objects are usually correlated, so object recognition based on independent decisions about components is inefficient, especially with squelching. Having to independently and rigorously detect the components makes efficiency inversely proportional to their number.  $\square$

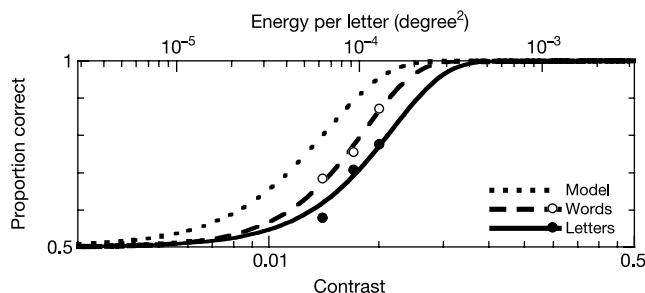
## Methods

### Testing

In each trial the signal was one of 26 possible letters (a ... z), or 26 possible words. The word list for each length (2–16 letters), for example 'of, to, in, ... , pa' for 2-letter words or 'responsibilities, misunderstandings, characterization, ..., aristocratically' for 16-letter words, consisted of the 26 most common words of that length, excluding words that are normally capitalized (such as 'American'), words containing punctuation, and nearly identical words<sup>24</sup>. The signal—a letter or word—was briefly presented (200 ms) either on a blank 50 cd m<sup>-2</sup> gamma-corrected screen (for human observers) or in gaussian noise (for both human and ideal observers). The blank screen and the noise had the same mean luminance (unlike the two rows of Fig. 1b). The task was to identify the signal by choosing one of the 26 signals, which were displayed for the human observer on an immediately-following response screen. (Increasing the viewing time, as in the printed demonstrations in Fig. 1, does not affect identification on the static noise background, but does aid identification on a blank background, because detection is then limited by the observer's intrinsic visual noise<sup>18</sup>, which is dynamic. See also ref. 15.) The text was rendered at 29 point in an off-screen image using the uniformly spaced TrueType Courier font (Fig. 1b), except for Figs 1a and 2, for which we used the proportionally spaced PostScript Adobe Bookman font. Efficiency of letter identification is slightly higher using Courier than using Bookman<sup>21</sup>. In the noise conditions, independent zero-mean gaussian samples, with standard deviation equal to 25% of the mean luminance, were added to the pixels of the off-screen image. The off-screen image was then doubled in size horizontally and vertically by pixel replication, and copied to the screen. The screen displayed 31.3 pixels per degree at the 0.6 m viewing distance. The power spectral density of the noise was  $N = 10^{-3.59} \text{ deg}^{-2}$ . The typographic x-height of the displayed text was 0.83 deg (Courier) or 0.89 deg (Bookman).

### Modelling

The results of the simulations at three contrasts were fit with a Weibull psychometric function, which is displayed as the dotted curve. The 26  $\times$  26 table of probabilities of each one-letter response to each one-letter signal is the observer's letter-confusion matrix. The simulation used 3,000-trial letter-confusion matrices measured for each observer at each of three contrasts in noise. (For the ideal they were 100,000-trial matrices at four contrasts.) To prevent bias due to correlated sampling errors between the confusion matrices used to simulate the first and second stages of the model, each stage used an



**Figure 4** The word superiority effect. Proportion correct as a function of contrast and average energy per letter, for letters and words on a blank background. The dotted line represents the upper bound for strictly letter-based word identification, based on the observer's measured confusion probabilities at three contrasts. Unlike Fig. 3, only two response alternatives, differing by only one letter, were offered on each trial, for example, 'coin' versus 'join' or 'c' versus 'j'. The word superiority is slight, no greater than can be accounted for by strictly letter-based identification.



independent 1,500- (or 50,000-) trial subset of the empirical letter-confusion counts. The first stage receives a letter  $i_1$  or a word  $i_1, i_2, i_3, i_4, i_5$  and independently emits a letter  $j_1$  or a letter at each position  $j_1, j_2, j_3, j_4, j_5$  with each letter probability specified by the  $26 \times 26$  confusion matrix  $P(j | i)$ . The best-possible second stage chooses the most probable 5-letter word  $i_1, i_2, i_3, i_4, i_5$  given the independent first-stage letter identifications  $j_1, j_2, j_3, j_4, j_5$ ; that is, it maximizes the posterior probability

$$P(i_1, i_2, i_3, i_4, i_5 | j_1, j_2, j_3, j_4, j_5) = \frac{P(i_1, i_2, i_3, i_4, i_5) P(j_1 | i_1) P(j_2 | i_2) P(j_3 | i_3) P(j_4 | i_4) P(j_5 | i_5)}{\sum_{i_1, i_2, i_3, i_4, i_5} P(i_1, i_2, i_3, i_4, i_5) P(j_1 | i_1) P(j_2 | i_2) P(j_3 | i_3) P(j_4 | i_4) P(j_5 | i_5)}$$

where  $P(i_1, i_2, i_3, i_4, i_5)$  is the prior probability of the word  $i_1, i_2, i_3, i_4, i_5$ . Concerned that the resulting 1,500 (3,000/2) trials per confusion matrix might not be enough, we simulated the two-stage model based on 100, 1,000, 3,000, 33,000 and 100,000 trials of the ideal observer. This revealed that the proportion correct is robust, only a few per cent lower for simulations using confusion matrices based on 3,000/2 rather than on 100,000/2 trials. Thus, our dotted curves in Figs 3 and 4 slightly underestimate the best-possible strictly letter-based performance, making our conclusion slightly more secure: human performance never clears the bar.

## Word superiority

Following Reicher's<sup>10</sup> elegant design, we began with a list of 288 pairs of 4-letter words that differed by only one letter within each pair. Equal numbers of words differed at each of the four letter positions. Our list was derived from that of ref. 30, replacing 18 obscure words (such as boll, lave, wile) by more common ones (ball, lake, mile). The observer was EG. In the word-identification task the observer was shown a low-contrast word, randomly selected from the list. Unlike our previous experiments, the response screen merely asked the observer to select between the correct word and its mate, which differed in only one letter position, as in 'c<sub>o</sub>i<sub>n</sub>' versus 'j<sub>o</sub>i<sub>n</sub>'. The letter-identification task drew from the same word list, but left blank all but the differing letters in each pair ('c' versus 'j') in both the stimulus and response screens. The ideal observer performs identically on both tasks, because the non-differing letters are irrelevant to the choice, so the human observer's word superiority implies that here the human is identifying words slightly more efficiently than letters. Thus, the word superiority effect is aptly named, but, as explained in the text, is consistent with strictly letter-based word identification.

Received 30 December 2002; accepted 21 February 2003; doi:10.1038/nature01516.

- Robson, J. G. & Graham, N. Probability summation and regional variation in contrast sensitivity across the visual field. *Vision Res.* **21**, 409–418 (1981).
- Diamond, R. & Carey, S. Why faces are and are not special: an effect of expertise. *J. Exp. Psychol. Gen.* **115**, 107–117 (1986).
- Gauthier, I., Skudlarski, P., Gore, J. C. & Anderson, A. W. Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neurosci.* **3**, 191–197 (2000).
- Treisman, A. & Schmidt, H. Illusory conjunctions in the perception of objects. *Cogn. Psychol.* **14**, 107–141 (1982).
- Biederman, I. Recognition-by-components: a theory of human image understanding. *Psychol. Rev.* **94**, 115–147 (1987).
- Tarr, M. J. & Buelthoff, H. H. Image-based object recognition in man, monkey and machine. *Cognition* **67**, 1–20 (1998).
- Pelli, D. G., Palomares, M. & Majaj, N. J. Crowding is unlike ordinary masking: distinguishing feature detection and integration. *J. Vision* [online] (<http://www.journalofvision.org/>) (in the press).
- Cattell, J. M. The time taken up by cerebral operations. *Mind* **11**, 220–242 (1886).
- Huey, E. B. *The Psychology and Pedagogy of Reading* (Macmillan, New York, 1908).
- Reicher, G. M. Perceptual recognition as a function of the meaningfulness of stimulus material. *J. Exp. Psychol.* **81**, 275–280 (1969).
- Wheeler, D. D. Processes in word recognition. *Cogn. Psychol.* **1**, 59–85 (1970).
- Johnston, J. C. & McClelland, J. L. Perception of letters in words: seek not and ye shall find. *Science* **184**, 1192–1194 (1974).
- Johnston, J. C. & McClelland, J. L. Experimental tests of a hierarchical model of word identification. *J. Verbal Learn. Verbal Behav.* **19**, 503–524 (1980).
- McClelland, J. L. & Rumelhart, D. E. An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychol. Rev.* **88**, 375–407 (1981).
- Prinzmetal, W. & Silvers, B. The word without the tachistoscope. *Percept. Psychophys.* **55**, 296–312 (1994).
- Jordan, T. R. & Bevan, K. M. Position-specific masking and the word-letter phenomenon: reexamining the evidence from the Reicher-Wheeler paradigm. *J. Exp. Psychol. Hum. Percept. Perform.* **22**, 1416–1433 (1996).
- Peterson, W. W., Birdsall, T. G. & Fox, W. C. Theory of signal detectability. *IRE Trans. Inf. Theory* **4**, 171–212 (1954).
- Pelli, D. G. & Farell, B. Why use noise? *J. Opt. Soc. Am. A* **16**, 647–653 (1999).
- Tanner, W. P. Jr & Birdsall, T. G. Definitions of  $d'$  and  $\eta$  as psychophysical measures. *J. Acoust. Soc. Am.* **30**, 922–928 (1958).
- Legge, G. E., Pelli, D. G., Rubin, G. S. & Schleske, M. M. Psychophysics of reading—I. Normal vision. *Vision Res.* **25**, 239–252 (1985).
- Pelli, D. G., Burns, C. W., Farell, B. & Moore, D. C. Identifying letters. *Vision Res.* (in the press).
- Campbell, F. W. & Robson, J. G. Application of Fourier analysis to the visibility of gratings. *J. Physiol. (Lond.)* **197**, 551–566 (1968).
- Watson, A. B. & Robson, J. G. Discrimination at threshold: labelled detectors in human vision. *Vision Res.* **21**, 1115–1122 (1981).
- Kucera, H. & Francis, W. N. *Computational Analysis of Present-Day American English* (Brown Univ. Press, Providence, 1967).
- Hadley, J. A. & Healy, A. F. When are reading units larger than the letter? Refinement of the Unitization Reading Model. *J. Exp. Psychol. Learn. Mem. Cogn.* **17**, 1062–1073 (1991).

- Barlow, H. B. Summation and inhibition in the frog's retina. *J. Physiol. (Lond.)* **119**, 69–88 (1953).
- Solomon, J. A. & Pelli, D. G. The visual filter mediating letter identification. *Nature* **369**, 395–397 (1994).
- Majaj, N. J., Liang, Y. X., Martelli, M., Berger, T. D. & Pelli, D. G. The channel for reading. *J. Vision* [online] **3** (<http://www.journalofvision.org/>) (2003).
- Selfridge, O. Pandemonium: a paradigm for learning. *Symposium on the Mechanization of Thought Processes* 513–526 (HM Stationery Office, London, 1959).
- Johnston, J. C. A test of the sophisticated guessing theory of word perception. *Cogn. Psychol.* **10**, 123–153 (1978).

Supplementary Information accompanies the paper on [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** Thanks to our many friends and colleagues who provided helpful comments, especially W. S. Geisler and R. F. Murray, who suggested that psychometric steepness may help to explain the reciprocal relation between efficiency and word length, J. M. Radner, who helped us say what we meant, I. Gauthier, D. J. Heeger, J. C. Johnston, M. S. Landy, G. E. Legge, J. M. Loomis, G. L. Murphy, R. E. Nixon, W. P. Prinzmetal and E. E. Smith. This work was supported by National Eye Institute grants to D.G.P. and B.F. D.C.M. was a Syracuse University undergraduate when she ran these experiments.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to D.G.P. ([denis.pelli@nyu.edu](mailto:denis.pelli@nyu.edu)).

## Control of dynamic CFTR selectivity by glutamate and ATP in epithelial cells

M. M. Reddy\* & P. M. Quinton\*†

\* Department of Pediatrics, UCSD School of Medicine, University of California, San Diego, La Jolla, California 92093-0831, USA

† Division of Biomedical Sciences, University of California, Riverside, California 92521, USA

Cystic fibrosis is caused by mutations in cystic fibrosis transmembrane conductance regulator (CFTR), an anion channel<sup>1</sup>. Phosphorylation and ATP hydrolysis are generally believed to be indispensable for activating CFTR<sup>2</sup>. Here we report phosphorylation- and ATP-independent activation of CFTR by cytoplasmic glutamate that exclusively elicits  $\text{Cl}^-$ , but not  $\text{HCO}_3^-$ , conductance in the human sweat duct. We also report that the anion selectivity of glutamate-activated CFTR is not intrinsically fixed, but can undergo a dynamic shift to conduct  $\text{HCO}_3^-$  by a process involving ATP hydrolysis. Duct cells from patients with  $\Delta\text{F508}$  mutant CFTR showed no glutamate/ATP activated  $\text{Cl}^-$  or  $\text{HCO}_3^-$  conductance. In contrast, duct cells from heterozygous patients with R117H/ $\Delta\text{F508}$  mutant CFTR also lost most of the  $\text{Cl}^-$  conductance, yet retained significant  $\text{HCO}_3^-$  conductance. Hence, not only does glutamate control neuronal ion channels, as is well known, but it can also regulate anion conductance and selectivity of CFTR in native epithelial cells. The loss of this uniquely regulated  $\text{HCO}_3^-$  conductance is most probably responsible for the more severe forms of cystic fibrosis pathology.

The molecular structure of CFTR is much more complicated than most ion channels. CFTR combines a regulatory domain consisting of numerous phosphorylation sites with two nucleotide-binding domains capable of ATP hydrolysis<sup>3</sup> flanked by six transmembrane domains on each side<sup>4</sup>. Consistent with this structure, a consensus has evolved that the phosphorylation by protein kinase A and hydrolysis of ATP are essential for activating CFTR  $\text{Cl}^-$  channel<sup>2</sup>. However, several discordant observations raised questions as to whether these requirements are absolute. We previously reported that CFTR  $\text{Cl}^-$  channels in the human sweat ducts are constitutively

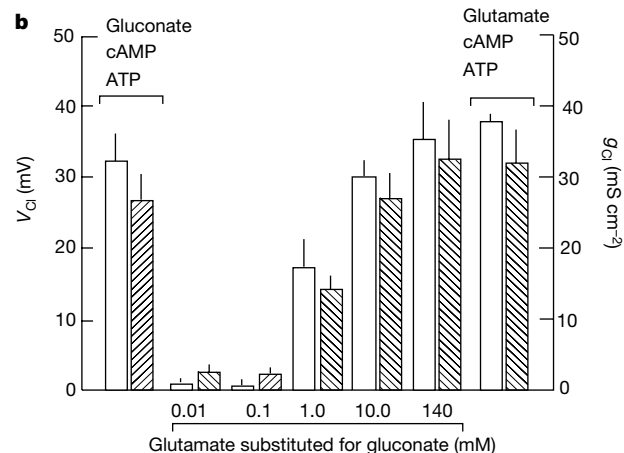
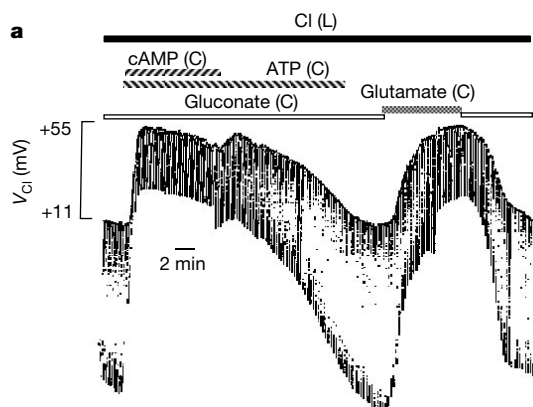
open and seem to be independent of regulation by intracellular cyclic AMP and protein kinase A phosphorylation activity<sup>5,6</sup>. In exploring alternative regulatory mechanisms, we focused on glutamate because of its predominant role in regulating neuronal ion channels and because glutamate receptors are widely distributed in epithelial cells even though the functional significance of these receptors is almost unknown<sup>7,8</sup>. We permeabilized microperfused sweat duct basolaterally with  $\alpha$ -toxin from *Staphylococcus aureus* and used the CFTR-rich<sup>6,9</sup> apical membrane to assay for changes in CFTR  $\text{Cl}^-$  conductance (CFTR- $g_{\text{Cl}}$ )<sup>5,10</sup>. Permeabilization of the basolateral membrane permitted control of the cytosolic composition of small electrolytes, nucleotides, amino acids and small organic anions by design<sup>11</sup>.

Figure 1 shows that cytoplasmic glutamate activates CFTR- $g_{\text{Cl}}$  in the absence of cAMP and ATP. Because glutamate is a potential source of energy by means of deamination and entry into the citric acid cycle, we first examined whether glutamate might be serving as a substrate for the endogenous metabolic production of ATP and cAMP<sup>12</sup> and might thereby account for its apparent stimulatory activity. We concluded that this was unlikely because amino acids (aspartate, taurine and leucine) (140 mM), organic anions (gluconate, cyclamate and glucuronate) (140 mM),  $\text{SO}_4^-$  (75 mM), glucose (10 mM), or most compellingly, ATP (5 mM), added to the cytosol in the same manner as glutamate did not activate CFTR- $g_{\text{Cl}}$  (data not shown). If ATP were being produced from glutamate, CFTR- $g_{\text{Cl}}$  should have been activated at least when 5 mM ATP was added. Similarly, it is not likely that glutamate elevated cAMP concentrations because forskolin plus 3-isobutyl-1-methyl xanthine, a treatment that almost universally elevates cAMP concentrations, did not activate CFTR<sup>5</sup>. Significantly, the non-selective kinase inhibitor staurosporine did not block the activation by glutamate (results not shown). These results indicate that glutamate activation of CFTR- $g_{\text{Cl}}$  did not require ATP or kinase phosphorylation.

Glutamate is effective at physiological concentrations. We estimated a  $K_m$  of  $\sim 3$  mM for the activation of CFTR by glutamate, which seems consistent with the reported cytosolic glutamate concentration ( $\sim 10$  mM)<sup>13,14</sup>. Only the L-isomer, characteristic of vertebrates<sup>15</sup>, was effective. Preliminary results indicated that

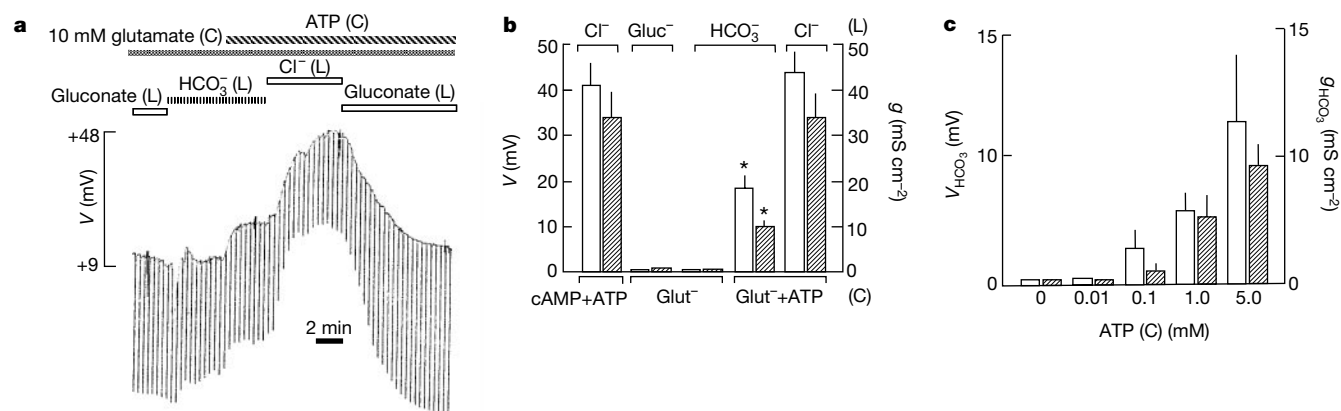
two closely related metabolites of glutamate ( $\alpha$ -ketoglutarate and glutamine) also induced a qualitatively similar activation of CFTR (probably owing to the secondary generation of glutamate by these native epithelial cells). Although the best-known receptor-mediated actions of glutamate are initiated from the extracellular surface<sup>7,8,15,16</sup>, it had no effect when applied to the luminal surface. However, glutamate has recently been shown to act as an intracellular messenger mediating insulin secretion by pancreatic  $\beta$ -cells<sup>15–17</sup>. Nothing is known of glutamate receptors in the sweat duct, but preliminary evidence reveals that sweat duct cells express mRNA for at least two metabotropic glutamate receptors (no. 3 and no. 4; results not shown).

It is generally accepted that ion selectivity is a fixed property inherent in channel activity and therefore that it is often used to define and distinguish ion channels. With the possible exception of a suggestion of state-dependent selectivity of certain cation channels in excitable cells<sup>18,19</sup> and ATP-dependent asymmetry of permeation of large organic anions through CFTR in heterologous expression systems<sup>20</sup>, we are unaware of any example of a ligand, or stimulation-coupled dynamic change, in the selectivity of an epithelial ion channel. However, we now see that CFTR is permeable to  $\text{Cl}^-$  and impermeable to  $\text{HCO}_3^-$  when activated by glutamate alone (or in combination with cAMP; results not shown), but Fig. 2 shows that adding ATP to glutamate in the cytoplasm induced a CFTR- $\text{HCO}_3^-$  conductance (CFTR- $g_{\text{HCO}_3^-}$ ) without affecting the magnitude of the CFTR- $g_{\text{Cl}}$ . This is direct evidence that the selectivity ( $g_{\text{HCO}_3^-}$  relative to  $g_{\text{Cl}}$ ) of an anion channel in a native epithelium can be altered by a physiological mediator. It is likely that not only ATP binding but also ATP hydrolysis is required to activate  $g_{\text{HCO}_3^-}$  because the non-hydrolysable ATP analogue  $\beta$ - $\gamma$ -imidoadenosine 5'-phosphate (AMP-PNP) failed to induce  $g_{\text{HCO}_3^-}$  (Fig. 3). However, like glutamate-activated  $g_{\text{Cl}}$ , glutamate/ATP-activated  $g_{\text{HCO}_3^-}$  is apparently not kinase dependent, for two reasons: first, the estimated  $K_m$  ( $\sim 2$  mM) for ATP activation of CFTR- $g_{\text{HCO}_3^-}$  (Fig. 2) is at least an order of magnitude higher than the  $K_m$  for ATP in most kinase-mediated phosphorylation reactions<sup>2</sup>; and second, glutamate/ATP activation of CFTR- $g_{\text{HCO}_3^-}$  was insensitive to staurosporine, a highly non-selective inhibitor of kinases in general (Fig. 3).



**Figure 1** Effect of cytoplasmic glutamate on CFTR- $g_{\text{Cl}}$ . **a**, Activation of CFTR- $g_{\text{Cl}}$  by glutamate. CFTR- $g_{\text{Cl}}$  was first deactivated by removing cAMP (10  $\mu\text{M}$ ) and ATP (5 mM) from the cytoplasmic bath. Glutamate substituted for equimolar gluconate in the cell (C) reactivated CFTR- $g_{\text{Cl}}$ . The top of the tracing records the transmembrane potential of the duct, in this case the  $\text{Cl}^-$  diffusion potential ( $V_{\text{Cl}}$ ). The downward deflections are due to transmembrane constant-current pulses (50 nA) and are inversely related to the membrane conductance (see Methods). **b**, Histogram of glutamate dose response for CFTR- $g_{\text{Cl}}$ . Significant activation of CFTR- $g_{\text{Cl}}$  occurred in the physiological range of

glutamate concentrations (1–10 mM). The magnitude of CFTR- $g_{\text{Cl}}$  activation was almost the same as that induced by cAMP plus ATP. No additional activation of CFTR- $g_{\text{Cl}}$  was observed when cAMP and ATP was added together to glutamate. Open bars, lumen-to-cell  $\text{Cl}^-$  diffusion potentials (left ordinate); hatched bars, CFTR- $g_{\text{Cl}}$  (right ordinate). Perfusion conditions: 150 mM KCl in the lumen (L)/140 mM total concentration of potassium gluconate plus potassium glutamate in the cell (C). Values are mean  $\pm$  s.e.m. for a minimum of three ducts from three normal subjects.



**Figure 2** Effect of ATP on the HCO<sub>3</sub><sup>-</sup> selectivity of glutamate-activated CFTR.

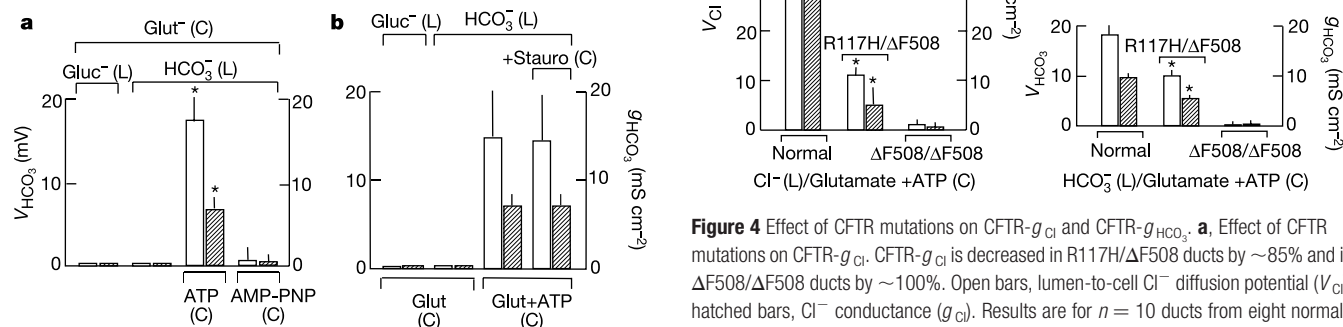
**a**, Experiment showing a lack of HCO<sub>3</sub><sup>-</sup> permeability from lumen to cell in the presence of 10 mM glutamate alone, as indicated by the almost complete absence of change in HCO<sub>3</sub><sup>-</sup> diffusion potential ( $V_{\text{HCO}_3}$ ). Adding 5 mM ATP in the presence of glutamate activated  $g_{\text{HCO}_3}$ , as indicated by significant increases in lumen positive HCO<sub>3</sub><sup>-</sup> diffusion potential and conductance. The trace was formed as in Fig. 1a. **b**, Histogram of effect of ATP on Cl<sup>-</sup>/HCO<sub>3</sub><sup>-</sup> selectivity of CFTR. Glutamate-activated CFTR without ATP was exclusively permeable to Cl<sup>-</sup> and not to HCO<sub>3</sub><sup>-</sup>, glutamate (Glut<sup>-</sup>) or gluconate (Gluc<sup>-</sup>). Open bars, lumen-to-cell diffusion potentials ( $V$ ) for the respective anions, that is, either 150 mM

gluconate, HCO<sub>3</sub><sup>-</sup> or Cl<sup>-</sup> in the lumen (L) against 140 mM total concentration of potassium gluconate plus potassium glutamate in the cell (C). Hatched bars, electrical conductance ( $g$ ) for the respective anions in the lumen. Results are for  $n = 6$  ducts from three normal subjects; asterisk,  $P > 0.01$  compared with  $g_{\text{HCO}_3}$  before stimulation with glutamate without ATP. **c**, Histogram of ATP dose response of CFTR- $g_{\text{HCO}_3}$ . These results show that CFTR- $g_{\text{HCO}_3}$  is activated in the physiological range of ATP concentrations (~5 mM). Open bars, lumen-to-cell HCO<sub>3</sub><sup>-</sup> diffusion potential ( $V_{\text{HCO}_3}$ ) generated by 150 mM KHCO<sub>3</sub> (L) against 140 mM total concentration of potassium gluconate plus potassium glutamate in the cell (C). Hatched bars,  $g_{\text{HCO}_3}$ . Results are for  $n = 7$  ducts from three normal subjects.

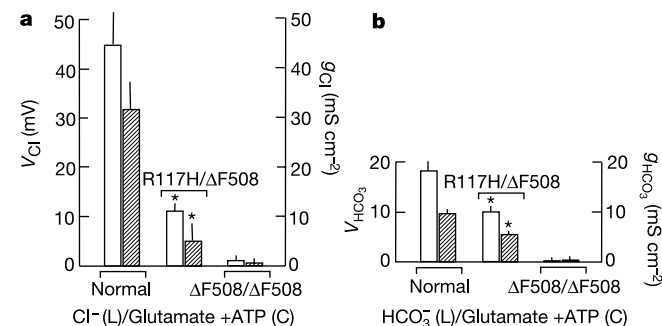
Thus, energy from ATP hydrolysis seems to be required for a conformational change in CFTR structure to confer HCO<sub>3</sub><sup>-</sup> permeability on an otherwise passive, ATP-independent Cl<sup>-</sup>-conducting channel. However, we do not yet know whether the activation of CFTR by glutamate is direct or indirect through an intracellular regulatory cascade.

We examined whether glutamate was activating an anion channel other than CFTR. The following evidence argues that this was not the case. First, activation by glutamate was neither synergistic nor additive when cAMP and ATP were present together. The magnitude of the maximum glutamate-activated Cl<sup>-</sup> conductance was almost the same as that maximally activated with cAMP plus ATP or

with all three agonists combined (Fig. 1), indicating that only one type of channel responded. Second, glutamate activated Cl<sup>-</sup> conductance in the complete absence of Ca<sup>2+</sup>, indicating that Ca<sup>2+</sup>-activated channels are probably not involved (not shown). Third, the observations that CFTR binds ATP<sup>21</sup> and is the only anion channel known to hydrolyse ATP are consistent with CFTR's being the HCO<sub>3</sub><sup>-</sup>-conductive pathway. Last, CFTR is the only known anion channel in the apical membrane of duct cells<sup>6,10,22</sup>.



**Figure 3** Effect of ATP and AMP-PNP on CFTR- $g_{\text{HCO}_3}$ . **a**, ATP, but not AMP-PNP, activates CFTR- $g_{\text{HCO}_3}$  in the presence of glutamate in the cell. Open bars, lumen-to-cell HCO<sub>3</sub><sup>-</sup> diffusion potential ( $V_{\text{HCO}_3}$ ) generated by 150 mM KHCO<sub>3</sub> (L)/140 mM total concentration of potassium gluconate plus potassium glutamate in the cell (C). Hatched bars,  $g_{\text{HCO}_3}$ . Glut<sup>-</sup>, glutamate; Gluc<sup>-</sup>, gluconate. Results are for  $n = 7$  ducts from three normal subjects; asterisk, significantly different from the controls without ATP and in the presence of AMP-PNP,  $P < 0.01$ . **b**, This figure shows the lack of effect of inhibiting the endogenous kinases with the non-selective kinase inhibitor staurosporine (Stauro) (1  $\mu\text{M}$ ) on glutamate/ATP activated CFTR- $g_{\text{HCO}_3}$ . Open and hatched bars have the same meanings as in **a**. Results are for  $n = 3$  ducts from three normal subjects.



**Figure 4** Effect of CFTR mutations on CFTR- $g_{\text{Cl}}$  and CFTR- $g_{\text{HCO}_3}$ . **a**, Effect of CFTR mutations on CFTR- $g_{\text{Cl}}$ . CFTR- $g_{\text{Cl}}$  is decreased in R117H/ΔF508 ducts by ~85% and in ΔF508/ΔF508 ducts by ~100%. Open bars, lumen-to-cell Cl<sup>-</sup> diffusion potential ( $V_{\text{Cl}}$ ); hatched bars, Cl<sup>-</sup> conductance ( $g_{\text{Cl}}$ ). Results are for  $n = 10$  ducts from eight normal subjects,  $n = 9$  ducts from two R117H/ΔF508 cystic fibrosis patients and  $n = 8$  ducts from three ΔF508/ΔF508 cystic fibrosis patients. Asterisk, significantly different as compared to both wild type and homozygous ΔF508 ducts,  $P > 0.01$ . **b**, Effect of R117H and ΔF508 CFTR mutations on glutamate/ATP-activated CFTR- $g_{\text{HCO}_3}$ . The ducts from ΔF508 homozygous cystic fibrosis patients are almost impermeable to HCO<sub>3</sub><sup>-</sup>. The ducts from R117H/ΔF508 cystic fibrosis patients retained ~50% of CFTR- $g_{\text{HCO}_3}$  compared with normal controls. Open bars, lumen-to-cell HCO<sub>3</sub><sup>-</sup> diffusion potential ( $V_{\text{HCO}_3}$ ); hatched bars, lumen-to-cell HCO<sub>3</sub><sup>-</sup> conductance ( $g_{\text{HCO}_3}$ ). Results are for  $n = 10$  ducts from eight normal subjects,  $n = 9$  ducts from two R117H/ΔF508 cystic fibrosis patients, and  $n = 8$  ducts from three homozygous ΔF508 cystic fibrosis subjects. Asterisk, significantly different as compared to both wild type and homozygous ΔF508 ducts,  $P > 0.01$ .



Perhaps the most compelling evidence that glutamate does not activate another anion channel is the fact that glutamate had no effect on duct cells from patients homozygous for the  $\Delta F508$  mutation (in which Phe 508 has been deleted) of the gene encoding CFTR (Fig. 4). This mutation blocks the processing of CFTR to the plasma membrane so that there is almost no CFTR-mediated  $g_{Cl}$  in affected cells<sup>23</sup>. In contrast, the R117H mutant CFTR, in which Arg 117 has been changed to His, is processed to the membrane and retains partial  $Cl^-$  conductance<sup>24,25</sup>. If another anion channel were the target of glutamate activation, then glutamate should have activated similar  $Cl^-$  conductances in cells of all three genotypes. As shown in Fig. 4,  $\Delta F508/\Delta F508$  ducts showed no detectable  $g_{Cl}$  response to glutamate, whereas the R117H/ $\Delta F508$  heterozygous ducts showed a significantly decreased (compared with the wild type), but clearly present,  $Cl^-$  conductance. Equally compelling evidence is that only R117H/ $\Delta F508$  duct cells (but not  $\Delta F508/\Delta F508$  cystic fibrosis ducts) retained a significant CFTR- $g_{HCO_3^-}$ . Thus, in the R117H/ $\Delta F508$  ducts we observed a decrease of nearly 85% in CFTR- $g_{Cl}$  whereas the same ducts retained about 50% of CFTR- $g_{HCO_3^-}$  (Fig. 4) compared with the wild type. (Because the R117H mutation significantly decreases the  $g_{Cl}$  and open probability of the CFTR<sup>25</sup>, the fact that at least 50% of normal  $g_{HCO_3^-}$  seems to be present in the heterozygous ducts suggests that this mutation exhibits a  $g_{HCO_3^-}$  that is equal to or even larger than normal. However, because of the small number of measurements, this value remains subject to statistical and experimental variation.) Nevertheless, we cannot exclude the possibility, though we think it remote, that another anion channel with very similar properties might be completely dependent on CFTR.

Since the cloning of the gene encoding CFTR, it has been possible to associate certain cystic fibrosis mutations with different levels of organ involvement. The phenotypes of such mutations are broadly classified as 'mild', in which patients fare better and retain pancreatic digestive function (for example R117H/ $\Delta F508$ ), and as 'severe', in which patients fare worse and lack pancreatic function (for example  $\Delta F508/\Delta F508$ ). A defect in  $HCO_3^-$  transport is demonstrated most notably in the pancreas, where the failure to secrete  $HCO_3^-$  produces autolysis and complete destruction of the exocrine organ<sup>26</sup>. A role for CFTR in  $HCO_3^-$  transport is still not well defined<sup>27–29</sup>, but the glutamate/ATP-activated CFTR- $g_{HCO_3^-}$  might reflect a crucial role in cystic fibrosis pathology and indicates a molecular basis for the observation that mild mutations such as R117H spare enough CFTR- $g_{HCO_3^-}$  to preserve pancreatic function in cystic fibrosis patients whereas severe mutations like  $\Delta F508$  do not, either because of poor conductance or poor expression in the plasma membrane<sup>23,24,26</sup> (Fig. 4). In heterologous systems,  $Cl^-/HCO_3^-$  exchange was reported to be dependent on the CFTR mutation, and R117H retained substantial ability to support the anion exchange function<sup>29,30</sup>. These authors suggested that the severity of pancreatic disease in cystic fibrosis might be related to variable abilities of different CFTR mutants to support  $Cl^-/HCO_3^-$  exchange. The large change in  $g_{HCO_3^-}$  makes it unlikely that glutamate is acting on an exchanger in the present studies. It therefore remains to be seen whether the loss of the electrodiffusive properties of CFTR to  $HCO_3^-$  are more consequential in the pathogenesis of cystic fibrosis than is the loss of  $Cl^-/HCO_3^-$  exchange, or even of CFTR- $g_{Cl}$ . But the new results presented here reinforce the notion that optimal therapy will probably require enhancing  $HCO_3^-$  transport in affected tissues. □

## Methods

Sweat ducts were obtained from skin biopsies of volunteer subjects who had given informed consent. Sweat ducts were dissected and microperfused as described previously<sup>11</sup>. Segments of ducts usually longer than 1,000  $\mu m$  were microperfused with a double-barrelled luminal micropipette that served to perfuse and record transepithelial voltage through one barrel and to pass constant current pulses through the other. This arrangement permitted the estimation of the specific membrane conductance from the cable equation<sup>22</sup>. After confirming the integrity of the perfused tubule, we applied 1,000–5,000 units  $ml^{-1}$  of

$\alpha$ -toxin from *Staphylococcus aureus* to the bath solution, to permeabilize the basolateral membrane selectively<sup>11</sup>. This procedure leaves the epithelium with an intact, functional apical membrane and a non-selective basal membrane, permeable to molecules with a relative molecular mass of up to ~5,000 including cAMP, ATP, glutamate, and other organic and inorganic ions. We then determined the permeability of the apical membrane to  $HCO_3^-$  and  $Cl^-$  relative to the impermeant gluconate anion by measuring the trans-apical membrane diffusion potential generated by chemical gradients for these anions and the simultaneous changes in membrane conductance after the activation of CFTR with cAMP, ATP and/or glutamate in the cytoplasm.

Our general experimental protocol was first to inactivate CFTR and perfuse the cytosol with potassium gluconate while changing the composition of the luminal perfusate from gluconate to  $Cl^-$  to  $HCO_3^-$  (we used the potassium salt because the duct apical membrane is impermeable to  $K^+$  ions). The integrity of the tissue was first confirmed by a large increase in CFTR- $g_{Cl}$  after the application of cAMP and ATP, as shown in Fig. 1a. cAMP was first washed out while leaving ATP in the cytoplasm until CFTR- $g_{Cl}$  returned to the baseline, indicating dephosphorylation and complete deactivation of CFTR. ATP was then removed from the cytoplasm before testing the effect of cytoplasmic glutamate on CFTR- $g_{Cl}$  and CFTR- $g_{HCO_3^-}$  relative to the impermeant anion gluconate.

The luminal perfusion Ringer's solutions contained (in mM) NaCl, KCl or  $KHCO_3^-$  (150),  $K^+$  (5),  $PO_4^{3-}$  (3.5),  $MgSO_4$  (1.2),  $Ca^{2+}$  (1), at pH 7.4.  $Cl^-$ -free luminal Ringer solution was prepared by the complete replacement of  $Cl^-$  with the impermeant anion gluconate. The cytoplasmic/bath solution contained 145 mM  $K^+$ , 140 mM gluconate, 3.5 mM  $PO_4^{3-}$ , 1.2 mM  $MgSO_4$  and 260  $\mu M$   $Ca^{2+}$  buffered with 2.0 mM EGTA (Sigma) to 80 nM free  $Ca^{2+}$ , pH 6.8. ATP, potassium salt (0.01–5 mM), cAMP (0.1 mM), glutamate substituted for gluconate (0.01–140 mM), AMP-PNP (5 mM) and staurosporine (0.001 mM) were added to the cytoplasmic bath as needed. After permeabilization of the basolateral membrane of the duct, the lumen was perfused with KCl, potassium gluconate or  $KHCO_3^-$  (each at 140 mM). Statistical significance was determined with Student's *t*-test;  $P < 0.05$  was taken to be significantly different.

Received 11 March; accepted 29 April 2003; doi:10.1038/nature01694.

1. Quinton, P. M. Physiological basis of cystic fibrosis: A historic perspective. *Physiol. Rev.* **79**, S3–S22 (1999).
2. Anderson, M. P. et al. Nucleoside triphosphates are required to open the CFTR chloride channel. *Cell* **67**, 775–784 (1991).
3. Li, C. et al. ATPase activity of the cystic fibrosis transmembrane conductance regulator. *J. Biol. Chem.* **271**, 28463–28468 (1996).
4. Riordan, J. R. et al. Identification of the cystic fibrosis gene: Cloning and characterization of complementary DNA. *Science* **245**, 1066–1072 (1989).
5. Reddy, M. M. & Quinton, P. M. cAMP-independent phosphorylation activation of CFTR by G proteins in native human sweat duct. *Am. J. Physiol. Cell Physiol.* **280**, C604–C613 (2001).
6. Lew, T. S. & Krasne, S. The only anion-selective channel measured in intact sweat ducts has a unit conductance of 8 pS. *Pediatr. Pulmonol. Suppl.* **6**, 251 (1991).
7. Duan, S. & Cooke, I. Glutamate and GABA activate different receptors and  $Cl^-$  conductances in Crab peptide secretory neurons. *J. Neurophysiol.* **83**, 31–37 (2000).
8. Cavalheiro, E. & Olney, J. Glutamate antagonists: Deadly liaisons with cancer. *Proc. Natl Acad. Sci. USA* **98**, 5947–5948 (2001).
9. Kartner, N., Augustinas, O., Jensen, T. J., Naismith, A. L. & Riordan, J. R. Mislocalization of  $\Delta F508$  CFTR in cystic fibrosis sweat gland. *Nature Genet.* **1**, 321–327 (1992).
10. Reddy, M. M. et al. Failure of the cystic fibrosis transmembrane conductance regulator to conduct ATP. *Science* **271**, 1876–1879 (1996).
11. Quinton, P. M. & Reddy, M. M. Control of CFTR  $Cl^-$  conductance by ATP levels through non-hydrolytic binding. *Nature* **360**, 79–81 (1992).
12. Welbourne, T. & Nissim, I. Regulation of mitochondrial glutamine/glutamate metabolism by glutamate transport: studies with  $^{15}N$ . *Am. J. Physiol. Cell Physiol.* **280**, C1151–C1159 (2001).
13. Raj, D., Langford, M., Krueger, S., Shelton, M. & Welbourne, T. Regulatory responses to an oral D-glutamate load: formation of D-pyrrolidone carboxylic acid in humans. *Am. J. Physiol. Renal Physiol.* **280**, F214–F220 (2001).
14. Dall'Asta, V. et al. Amino acids are compatible osmolytes for volume recovery after hypertonic shrinkage in vascular endothelial cells. *Am. J. Physiol. Cell Physiol.* **276**, C865–C872 (1999).
15. Skerry, T. & Genever, P. Glutamate signaling in non-neuronal tissues. *Trends Pharmacol. Sci.* **22**, 174–181 (2001).
16. Conn, P. & Pin, J. Pharmacology and functions of metabotropic glutamate receptors. *Annu. Rev. Physiol.* **37**, 205–237 (1997).
17. Wollheim, C. Beta-cell mitochondria in the regulation of insulin secretion: a new culprit in Type II diabetes. *Diabetologia* **43**, 265–277 (2002).
18. Santana, L. F., Gomez, A. M. & Lederer, W. J.  $Ca^{2+}$  flux through promiscuous cardiac  $Na^+$  channels: slip-mode conductance. *Science* **279**, 1027–1033 (1998).
19. Khakh, B. S. Molecular physiology of P2X receptors and ATP signalling at synapses. *Nature Rev. Neurosci.* **2**, 165–174 (2001).
20. Linsdell, P. & Hanrahan, J. W. Adenosine triphosphate dependent asymmetry of anion permeation in the cystic fibrosis transmembrane conductance regulator. *J. Gen. Physiol.* **111**, 601–614 (1998).
21. Gunderson, K. L. & Kopito, R. R. Conformational states of CFTR associated with channel gating: The role of ATP binding and hydrolysis. *Cell* **82**, 231–239 (1995).
22. Reddy, M. M., Light, M. J. & Quinton, P. M. Activation of the epithelial  $Na^+$  channel (ENaC) requires CFTR  $Cl^-$  channel function. *Nature* **402**, 301–304 (1999).
23. Cheng, S. H. et al. Defective intracellular transport and processing of CFTR is the molecular basis of most cystic fibrosis. *Cell* **63**, 827–834 (1990).
24. Drumm, M. L. et al. Chloride conductance expressed by  $\Delta F508$  and other mutant CFTRs in *Xenopus* oocytes. *Science* **254**, 1797–1799 (1991).
25. Sheppard, D. N. et al. Mutations in CFTR associated with mild-disease-form  $Cl^-$  channels with altered pore properties. *Nature* **362**, 160–164 (1993).
26. Durie, P. R. & Forstner, G. G. in *Cystic Fibrosis in Adults* (eds Yankaskas, J. R. & Knowles, M. R.) 261–287 (Lippincott-Raven, Philadelphia, 1999).

27. Sohma, Y., Gray, M. A., Imai, Y. & Argent, B.  $\text{HCO}_3^-$  transport in a mathematical model of the pancreatic ductal epithelium. *J. Membr. Biol.* **176**, 77–100 (2000).
28. Poulsen, J. H., Fischer, H., Illek, B. & Machen, T. E. Bicarbonate conductance and pH regulatory capability of cystic fibrosis transmembrane conductance regulator. *Proc. Natl Acad. Sci. USA* **91**, 5340–5344 (1995).
29. Choi, J. Y. *et al.* Aberrant CFTR-dependent  $\text{HCO}_3^-$  transport in mutations associated with cystic fibrosis. *Nature* **410**, 94–97 (2001).
30. Ko, S. B. H. *et al.* A molecular mechanism for aberrant CFTR-dependent  $\text{HCO}_3^-$  transport in cystic fibrosis. *EMBO J.* **21**, 5662–5672 (2002).

**Acknowledgements** We thank K. Taylor and S. Madireddi for technical assistance, A. Ponce for reverse transcriptase polymerase chain reaction, and M. Pian, D. Conrad, C. Nagy, A. Wallace and M. Farrel for assistance in securing specimens from normal and cystic fibrosis patients. This work was funded by grants from the USPHS-NIH, the CF Foundation, the Olmsted Trust, the Gillette Co. and the Texaco Foundation.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to M.M.R. (mmr@ucsd.edu).

## Pathogen-induced systemic plant signal triggers DNA rearrangements

Igor Kovalchuk\*, Olga Kovalchuk\*, Véronique Kalck†, Vitaly Boyko†, Jody Filkowski\*, Manfred Heinlein† & Barbara Hohn†

\* Department of Biological Sciences, University of Lethbridge, Lethbridge, Alberta T1K 3M4, Canada

† Friedrich Miescher-Institut for Biomedical Research, PO Box 2543, CH-4002 Basel, Switzerland

Plant genome stability is known to be affected by various abiotic environmental conditions<sup>1–7</sup>, but little is known about the effect of pathogens. For example, exposure of maize plants to barley stripe mosaic virus seems to activate transposable elements<sup>8,9</sup> and to cause mutations in the non-infected progeny of infected plants<sup>10</sup>. The induction by barley stripe mosaic virus of an inherited effect may mean that the virus has a non-cell-autonomous influence on genome stability. Infection with *Peronospora parasitica* results in an increase in the frequency of somatic recombination in *Arabidopsis thaliana*<sup>11</sup>; however, it is unclear whether effects on recombination require the presence of the pathogen or represent a systemic plant response. It is also not clear whether the changes in the frequency of somatic recombination can be inherited. Here we report a threefold increase in homologous recombination frequency in both infected and non-infected tissue of tobacco plants infected with either tobacco mosaic virus<sup>12</sup> or oilseed rape mosaic virus<sup>13</sup>. These results indicate the existence of a systemic recombination signal that also results in an increased frequency of meiotic and/or inherited late somatic recombination.

To analyse genome stability we used two different reporter systems in tobacco plants: one based on recombination-dependent activation of a luciferase transgene, the other based on mutation and recombination of the endogenous *Sulfur* (*Su*) gene<sup>14,15</sup>. Of interest was not only to test the possible response of the plant genome to virus infection but also to investigate whether local virus infection could cause non-cell-autonomous signalling throughout the plant, resulting in genomic change in non-infected as well as infected tissues.

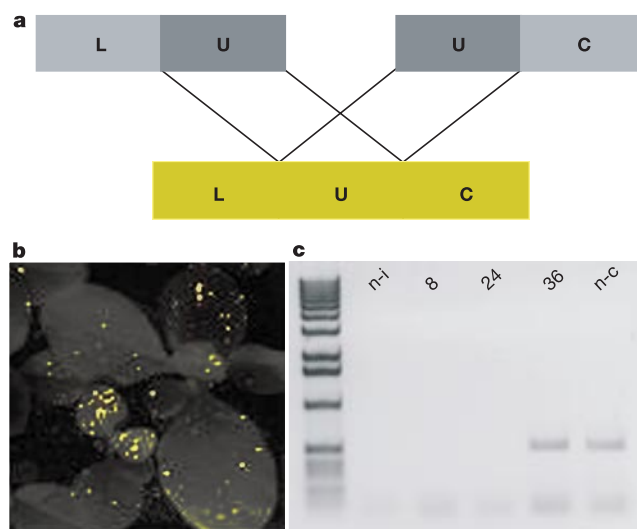
Homologous recombination was monitored in *Nicotiana tabacum* cv. Havana 425 plants transgenic for a recombination substrate consisting of two non-functional, overlapping copies of a luciferase gene (Fig. 1a). The homologous recombination repair of strand

breaks occurring in the regions of homology results in regeneration of the luciferase transgene, the activity of which becomes visible as bright spots on dark tissue (Fig. 1b).

To detect changes in the recombination frequency in response to a virus, we inoculated plants at 32 °C with tobacco mosaic virus (TMV)<sup>12</sup>. In *N. tabacum* cv. Havana 425, which carries the *N* resistance gene, infection causes a hypersensitive response, with formation of local lesions that restrict the virus to infected cells in the inoculated leaf. The *N*-mediated hypersensitive response, however, is sensitive to temperatures of 30 °C and higher. Under these conditions, the virus can spread systemically<sup>16</sup>. Seven days after inoculation, an average of  $28.5 \pm 7.2$  luciferase spots per plant could be counted, as compared with  $8.4 \pm 3.1$  spots in mock-inoculated plants, representing a 3.4-fold increase in recombination frequency in the inoculated leaves. An increase, albeit smaller (1.8-fold), in the recombination frequency in non-inoculated leaves was also observed.

To analyse whether the recombination-promoting signal is the spreading virus itself or a message triggered by the virus that can spread in advance of the virus to tissues distant from infected cells, we first examined the kinetics of virus movement. Single leaves of 50 plants were inoculated with TMV at 32 °C, and at five different times after inoculation the inoculated leaf was removed from 10 plants. The presence of virus in non-inoculated parts of the tobacco plants was analysed 7 d after inoculation by inspecting for disease symptoms and by analysing leaf extracts for the presence of viral RNA. Both tests clearly showed that the virus entered non-inoculated leaves between 24 and 36 h after inoculation (Fig. 1c). However, recombination events in non-inoculated leaves of the plants showed an increase in homologous recombination as early as 8 h after inoculation (Table 1). Thus, it seems that a signal leading to induction of homologous recombination can travel to new leaves faster than the virus itself can move.

To test whether such a 'recombination signal' can be also transmitted from plant to plant, we carried out grafting experiments



**Figure 1** The presence of TMV is not required for the increase in recombination frequency. **a**, Monitoring system for homologous recombination. L, U, C indicate the N-terminal, central and C-terminal parts of the luciferase gene, respectively. **b**, Recombination events visualized as bright spots on plants. **c**, The presence or absence of viral RNA was tested by RT-PCR on RNA prepared from the upper leaves of either non-inoculated (n-i) plants, or inoculated plants from which the inoculated leaves were cut at 8, 24, or 36 h after inoculation or were not cut at all (n-c). Typically, one leaf per treated plant was taken for analysis.

(Fig. 2a). Upper leaves from inoculated plants growing at 32 °C, from which the treated leaves had been removed 24 h after inoculation, were grafted onto healthy, non-infected transgenic plants. The grafted leaves did not contain virus, as tested by the combination of analyses described above. Before grafting, the top of the recipient root stock plant was excised to stimulate lateral shoot development. Homologous recombination was scored in newly emerged leaves 10–14 d after grafting.

A recombination-inducing signal was indeed detected: the recombination frequency was 2.3 times higher when plants were grafted with a 'signal-carrying' leaf than when plants were grafted with a leaf from a mock-inoculated plant ( $P < 0.01$ , 99% confidence interval,  $t$ -test  $1.4 \times 10^{-5}$ ;  $F$ -test  $5.3 \times 10^{-5}$ ; see Supplementary Table 1). Thus, the signal can be transmitted from plant to plant in the complete absence of the virus that originally incites its production.

To corroborate the data obtained with TMV as the virus pathogen, we analysed the effect of another virus, oilseed rape mosaic virus (ORMV), on somatic recombination frequency. A similar increase in the frequency of homologous recombination in new leaves of grafted plants was observed (Supplementary Table 2).

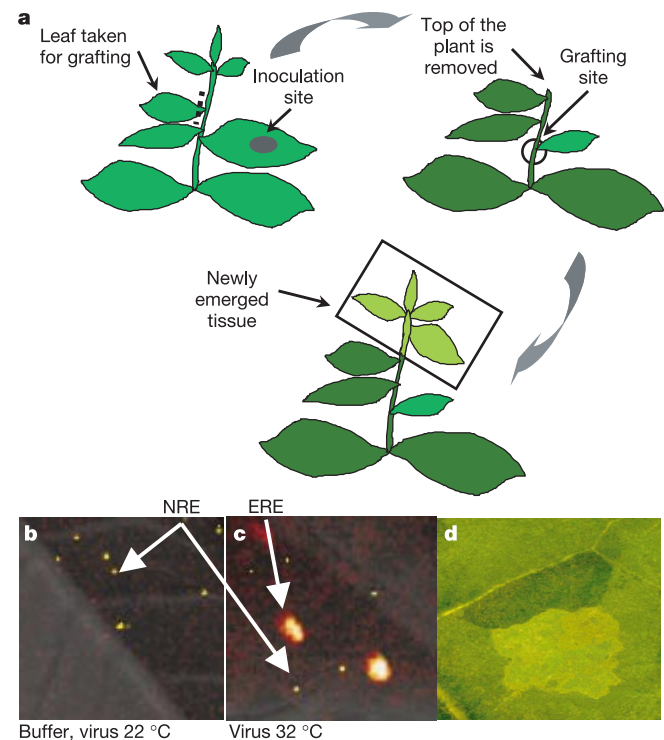
To verify that similar DNA rearrangements can also occur in an endogenous tobacco gene, we used *N. tabacum* cv. Xanthi NN plants (this cultivar also contains the *N* resistance gene) heterozygous for *Su*, a nuclear gene that affects chlorophyll content in leaves<sup>13</sup>. Various genetic events, including homologous recombination, can result in the spontaneous appearance of dark green (*Su/Su*), albino (*su/su*) or twin (*Su/Su* and *su/su*) sectors on pale green *Su/su* plants (ref. 15 and Fig. 2d). When non-infected leaves of ORMV-inoculated *Su/su* plants were grafted onto non-inoculated tester plants ( $P < 0.05$ , 95% confidence interval, 0.016  $t$ -test two-tailed paired; 0.061 Fisher test; see Supplementary Table 3).

Our results suggest the existence of a systemic recombination signal (SRS) that is elicited by virus infection and can move through the plant, triggering genomic change. It is possible that this SRS may be related to the so far chemically unidentified phloem mobile signal that is implicated in systemic acquired resistance (SAR)<sup>17,18</sup>. Indeed, the *N* gene required for SAR in response to TMV infection is present in our plants, although it is in an inactive state owing to high temperature. It is unlikely, however, that the systemic increase in recombination frequency triggered by local viral infection involves an *N*-mediated pathway because similar increases in recombination frequency on TMV infection were also measured in *N. tabacum* cv. Petit Havana SR1, which lacks an *N* gene.

Notably, similar increases in the number of recombination events were measured in SR1 plants at high (32 °C) and low (22 °C) temperature, indicating that the recombination increase in virus-inoculated plants does not depend on the post-inoculation temperature *per se* (Supplementary Information). By contrast, on inoculation of *N. tabacum* cv. Havana 425 plants at conditions permissive for *N* (22 °C), neither inoculated nor non-inoculated leaves showed an increase in homologous recombination frequency

(data not shown). Similarly, in grafting experiments, upper leaves from virus-inoculated plants grown at 22 °C could not trigger a recombination-promoting signal in newly emerging leaves of transgenic test plants (Supplementary Table 1). Thus, under conditions permissive for *N* activity, as well as for *N*-mediated hypersensitive response and SAR, systemic activation of recombination seems to be blocked. By contrast, the same conditions (22 °C) allowed the systemic activation of recombination in plants deficient for *N* (see above). Thus, we conclude that systemic activation of recombination is not dependent on the temperature, the *N* gene, or the SAR signal.

Whether genomic change is confined to homologous recombination, as measured in this study, or can be extended to other types of genome rearrangement needs further analysis. To analyse whether the SRS leads to inherited rearrangements in the *LUC* transgene and whether it influences meiotic recombination, we screened the progeny of grafted plants for those carrying a completely recombined *LUC* transgene (*LUC*<sup>+</sup>) (Supplementary Fig. 1). We collected the seeds of plants grafted with leaves from either infected or buffer-treated plants. We found 60 *LUC*<sup>+</sup> plants among 63,000 progeny of 'infected' plants ( $1.0 \times 10^{-3} \pm 0.5 \times 10^{-3}$ ) and 10 *LUC*<sup>+</sup> plants among 34,000 progeny of 'buffer-treated' plants ( $2.9 \times 10^{-4} \pm 3.2 \times 10^{-4}$ ). Thus, the SRS resulted in the recovery of 3.2 times more plants fully recombined in the transgene ( $P < 0.01$ , 99% confidence interval, 0.00023  $t$ -test two-tailed paired; 0.15 Fisher test). This could be a consequence of an



**Figure 2** Grafting experiments. **a**, Presentation of grafting experiments. Leaves for grafting were taken 24 h after inoculation. Plants were assayed 10–14 d after grafting. **b**, Newly emerged leaves (boxed in **a**) were taken for analysis. Fluorescent spots of 'normal' size, indicating normal recombination events (NRE), were observed in plants grafted with upper leaves of buffer-treated plants or plants inoculated with virus at 22 °C (see Supplementary Table 1). **c**, Fluorescent spots of different sizes, indicating normal recombination events or 'early' recombination events (ERE), were observed in plants grafted with upper leaves of plants inoculated with virus at 32 °C. The luciferase spots were considered as normal recombination events if they were less than 1 mm in diameter and early recombination events if they were larger than 1 mm in diameter (Supplementary Table 1). **d**, Twin (dark and albino) sector in *Su/su* plant leaf.

**Table 1** Recombination increase after inoculation

| Time (h) | Symptoms | Fold increase |
|----------|----------|---------------|
| 0        | —        | 1.0           |
| 8        | —        | 1.8–2.2       |
| 24       | —        | 2.0–3.0       |
| 36       | +        | 2.0–3.0       |
| Not cut  | +        | 2.0–3.0       |

Recombination increase (relative to control) analysed in non-inoculated leaves 7 d after inoculation. Single leaves of 50 plants were inoculated with TMV at 32 °C and the inoculated leaves were removed from 10 plants at different times after inoculation. 'Fold increase' represents the average fold increase of recombination frequency over the control value.



increased frequency of meiotic and/or inherited late somatic recombination.

It will be interesting to elucidate further the relationship of SRS with SAR signalling as well as with systemic wound signalling<sup>19</sup>, systemic post-transcriptional gene silencing<sup>20</sup> and systemic acquired acclimatization to light<sup>21</sup>. The genomic changes reported here might constitute an adaptive measure in response to biotic stress. Of special utility may be recombination events that lead to new specificities in pathogen resistance genes<sup>22</sup>. □

## Methods

### Generation of transgenic plants

The generation and analysis of luciferase recombination test plants has been described<sup>23</sup>.

### Plant inoculation

Single leaves of 6–8-week-old TMV-sensitive Havana 425 tobacco plants were rub-inoculated with 300 ng of full-length infectious TMV or ORMV RNA transcripts as described<sup>16</sup>; control plants were mock-inoculated with phosphate buffer. After treatment, plants were kept at 22 °C or 32 °C and recombination frequencies were scored 7 d after inoculation. In another experiment, single TMV-inoculated leaves of plants inoculated at 32 °C were excised and removed from the plants at times between 0 and 36 h (5 min, 8 h, 24 h, 36 h) or left on the plants. We detected disease symptoms on non-inoculated leaves 10 d after inoculation. Increase of recombination (ratio to control) was analysed in non-inoculated leaves 7 d after inoculation.

### Assay for infection

The presence or absence of the viral RNA was tested by polymerase chain reaction with reverse transcription (RT-PCR; primers: forward, 5'-CTGGTGAAGTATTTGTCTGA-3', and reverse, 5'-ACCCGCTGACATCTTCACAT-3') done on the remaining tissue 2 weeks after inoculation (Fig. 1c).

### Grafting experiments

We inoculated single leaves of several 10-week-old Havana 425 tobacco plants with TMV at 22 °C or 32 °C, and after 24 h we grafted upper, virus-free leaves from these plants onto 10-week-old healthy plants. The scion leaves were grafted onto the sides of the stem in place of the removed third leaf (Fig. 2a). After grafting, the plants were kept at either 22 °C or 32 °C. We measured the recombination frequency in the newly emerged leaves 10–14 d after grafting.

*Su/su* tobacco plants were infected with 300 ng of ORMV RNA. Grafting experiments similar to those done with TMV were conducted. We scored dark green and albino sectors on newly emerged tissue 1 month after grafting (Fig. 2d).

### Statistical treatment of the data

Averages and standard deviations were calculated in all experiments. The statistical significance of the experiments was confirmed by Student's *t*-test (two-tailed paired or non-paired) and the Fisher test.

Received 20 December 2002; accepted 14 April 2003; doi:10.1038/nature01683.

- Puchta, H., Swoboda, P. & Hohn, B. Induction of homologous DNA recombination in whole plants. *Plant J.* **7**, 203–210 (1995).
- Kovalchuk, L., Kovalchuk, O., Arkhipov, A. & Hohn, B. Transgenic plants are sensitive bioindicators of nuclear pollution caused by the Chernobyl accident. *Nature Biotechnol.* **16**, 1054–1057 (1998).
- Ries, G. *et al.* Elevated UV-B radiation reduces genome stability in plants. *Nature* **406**, 98–101 (2000).
- Ries, G., Buchholz, G., Frohnmeyer, H. & Hohn, B. UV-damage-mediated induction of homologous recombination in *Arabidopsis* is dependent on photosynthetically active radiation. *Proc. Natl Acad. Sci. USA* **97**, 13425–13429 (2000).
- Walbot, V. Reactivation of Mutator transposable elements of maize by ultraviolet light. *Mol. Gen. Genet.* **234**, 353–360 (1992).
- Walbot, V. UV-B damage amplified by transposons in maize. *Nature* **397**, 398–399 (1999).
- Kumar, A. & Bennetzen, J. L. Plant retrotransposons. *Annu. Rev. Genet.* **33**, 479–532 (1999).
- Dellaporta, S. *et al.* Endogenous transposable elements associated with virus infection in maize. *Cold Spring Harbor Symp. Quant. Biol.* **49**, 321–328 (1984).
- Johns, M. A., Mottinger, J. & Freeling, M. A low copy number, copia-like transposon in maize. *EMBO J.* **4**, 1093–1102 (1985).
- Brakke, M. Mutations, the aberrant ratio phenomenon, and virus infection of maize. *Annu. Rev. Phytopathol.* **22**, 77–94 (1984).
- Lucht, J. M. *et al.* Pathogen stress increases somatic recombination frequency in *Arabidopsis*. *Nature Genet.* **30**, 311–314 (2002).
- Harrison, B. D. & Wilson, T. M. A. Milestones in the research on tobacco mosaic virus. *Phil. Trans. R. Soc. Lond. B* **354**, 521–529 (1999).
- Aguilar, I., Sanchez, E., Martin, A., Martinez-Herrera, D. & Ponz, F. Nucleotide sequence of Chinese rape mosaic virus (oilseed rape mosaic virus), a crucifer tobamovirus infecting *Arabidopsis thaliana*. *Plant Mol. Biol.* **30**, 191–197 (1996).
- Burk, L. G. & Menger, H. A. A dominant *aurea* mutation in tobacco. *Tob. Sci.* **8**, 101–104 (1964).
- Shalev, G., Sitrit, Y., Avivi-Ragolski, N., Lichtenstein, C. & Levy, A. A. Stimulation of homologous recombination in plants by expression of the bacterial resolvase RuvC. *Proc. Natl Acad. Sci. USA* **96**, 7398–7402 (1999).
- Boyko, V., Ferralli, J., Ashby, J., Schellenbaum, P. & Heinlein, M. Function of microtubules in intercellular transport of plant virus RNA. *Nature Cell Biol.* **2**, 826–832 (2000).

- Dempsey, D. A., Shah, J. & Klessig, D. Salicylic acid and disease resistance in plants. *Crit. Rev. Plant Sci.* **18**, 547–575 (1999).
- Dong, X. Genetic dissection of systemic acquired resistance. *Curr. Opin. Plant Biol.* **4**, 309–314 (2001).
- Pearce, G., Strydom, D., Johnson, S. & Ryan, C. A polypeptide from tomato leaves induces wound-inducible proteinase inhibitor proteins. *Science* **253**, 895–897 (1991).
- Waterhouse, P., Wang, M.-B. & Finnegan, J. Role of short RNAs in gene silencing. *Trends Plant Sci.* **6**, 297–301 (2001).
- Karpinski, S. *et al.* Systemic signalling and acclimation in response to excess excitation energy in *Arabidopsis*. *Science* **284**, 654–657 (1999).
- Whitham, S. *et al.* The product of the tobacco mosaic virus resistance gene *N*: similarity to toll and the interleukin-1 receptor. *Cell* **78**, 1101–1115 (1994).
- Gorbunova, V. *et al.* A new hyperrecombinogenic mutant of *Nicotiana tabacum*. *Plant J.* **24**, 601–611 (2000).

Supplementary Information accompanies the paper on [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank S. van Eeden for technical assistance and H. Rothnie, K. Smith and E. Schultz for comments on the manuscript. The Novartis Research Foundation and Alberta Ingenuity Grant are acknowledged for financial support.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to I.K. (igor.kovalchuk@uleth.ca).

## Cloning of adiponectin receptors that mediate antidiabetic metabolic effects

Toshimasa Yamauchi\*†‡, Junji Kamon\*‡§, Yusuke Ito\*, Atsushi Tsuchida\*, Takehiko Yokomizo||, Shunbun Kita\*, Takuya Sugiyama¶, Makoto Miyagishi#, Kazuo Hara\*†, Masaki Tsunoda☆, Koji Murakami☆, Toshiaki Ohteki\*\*†, Shoko Uchida\*, Sato Takekawa\*, Hironori Waki\*†, Nelson H. Tsuno††, Yoichi Shibata††, Yasuo Terauchi\*†, Philippe Froguel‡‡, Kazuyuki Tobe\*†, Shigeo Koyasu\*†, Kazunari Taira#, Toshio Kitamura¶, Takao Shimizu||, Ryozi Nagai\* & Takashi Kadowaki\*†

\* Department of Internal Medicine, Graduate School of Medicine, University of Tokyo, Tokyo 113-8655, Japan

† CREST of Japan Science and Technology Corporation, 332-0012, Japan

§ Biological Research Laboratories, Nissan Chemical Industries, Saitama 349-0294, Japan

|| Department of Biochemistry and Molecular Biology, Faculty of Medicine, University of Tokyo, Tokyo 113-0033, CREST and PRESTO of JST, Japan

¶ Division of Hematopoietic Factors, Institute of Medical Science, University of Tokyo, Tokyo 108-8639, Japan

# Department of Chemistry and Biotechnology, School of Engineering, University of Tokyo, Tokyo 113-8656, and Gene Function Research Center, National Institute of AIST, Tsukuba 305-8562, Japan

☆ Central Research Laboratories, Kyorin Pharmaceutical, Tochigi 329-0114, Japan

\*\* Department of Microbiology and Immunology, Keio University School of Medicine, Tokyo 160-8582, Japan

†† Department of Transfusion Medicine, Graduate School of Medicine, University of Tokyo, Tokyo 113-8655, Japan

‡‡ Institute of Biology-CNRS, Pasteur Institute of Lille, UPRES A8090, 59000 Lille, France

‡ These authors contributed equally to this work

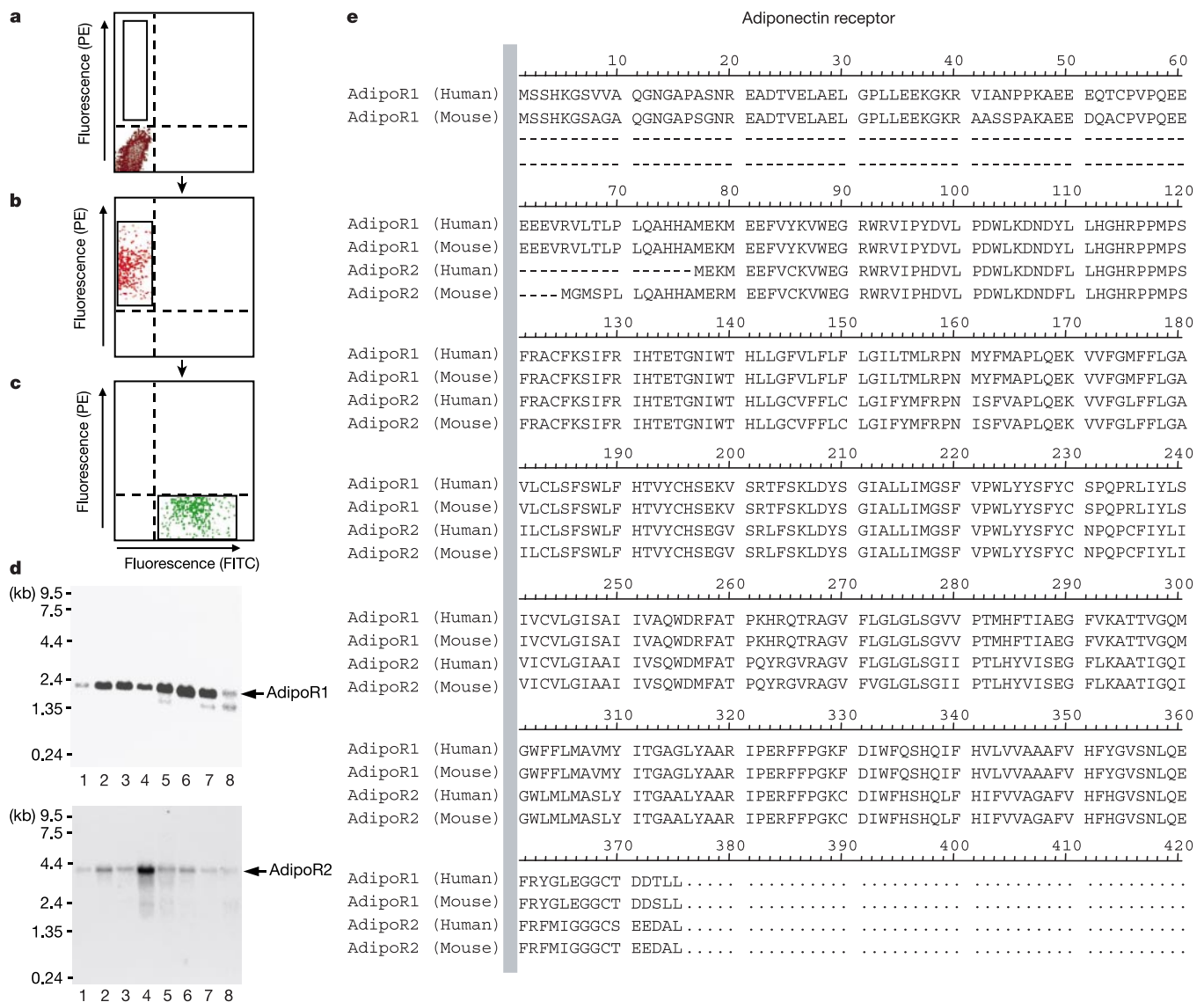
Adiponectin (also known as 30-kDa adipocyte complement-related protein; Acrp30)<sup>1–4</sup> is a hormone secreted by adipocytes that acts as an antidiabetic<sup>5–12</sup> and anti-atherogenic<sup>8,12,13</sup> adipokine. Levels of adiponectin in the blood are decreased under conditions of obesity, insulin resistance and type 2 diabetes<sup>2</sup>. Administration of adiponectin causes glucose-lowering effects

and ameliorates insulin resistance in mice<sup>5-7</sup>. Conversely, adiponectin-deficient mice exhibit insulin resistance and diabetes<sup>8,9</sup>. This insulin-sensitizing effect of adiponectin seems to be mediated by an increase in fatty-acid oxidation through activation of AMP kinase<sup>10,11</sup> and PPAR- $\alpha$ <sup>5,6,12</sup>. Here we report the cloning of complementary DNAs encoding adiponectin receptors 1 and 2 (AdipoR1 and AdipoR2) by expression cloning<sup>14-16</sup>. AdipoR1 is abundantly expressed in skeletal muscle, whereas AdipoR2 is predominantly expressed in the liver. These two adiponectin receptors are predicted to contain seven transmembrane domains, but to be structurally and functionally distinct from G-protein-coupled receptors<sup>17-19</sup>. Expression of AdipoR1/R2 or suppression of AdipoR1/R2 expression by small-interfering RNA<sup>20</sup> supports our conclusion that they serve as receptors for globular and full-length adiponectin, and that they mediate

increased AMP kinase<sup>10,11</sup> and PPAR- $\alpha$  ligand activities<sup>12</sup>, as well as fatty-acid oxidation and glucose uptake by adiponectin.

Globular adiponectin binds more avidly than full-length adiponectin in C2C12 myocytes and skeletal muscle membranes, but this pattern of binding is reversed in hepatocytes and liver membranes<sup>10</sup> (Supplementary Fig. 1a, b). We attempted to isolate cDNA for adiponectin receptors—which mediate antidiabetic effects—from Ba/F3 cells infected with a library of retrovirally expressed cDNA, derived from human skeletal muscle messenger RNA, by screening for globular adiponectin binding<sup>14</sup>.

Infected Ba/F3 cells were incubated with globular adiponectin with a red fluorescent probe and subjected to three rounds of sorting (Fig. 1a, b); the sorted cells were then incubated with globular adiponectin with a green fluorescent probe. As the cells that changed from red to green may have specific binding sites for



**Figure 1** Expression cloning of adiponectin receptors. **a–c**, Globular adiponectin binding to Ba/F3 cells transfected with a skeletal muscle cDNA library. **a**, Transfected cells were incubated with biotinylated globular adiponectin stained with a red fluorescent probe (phycoerythrin, PE). **b**, Cells before the third sorting. **c**, Cells incubated with globular adiponectin conjugated with fluorescein isothiocyanate (FITC) before the fourth sorting.

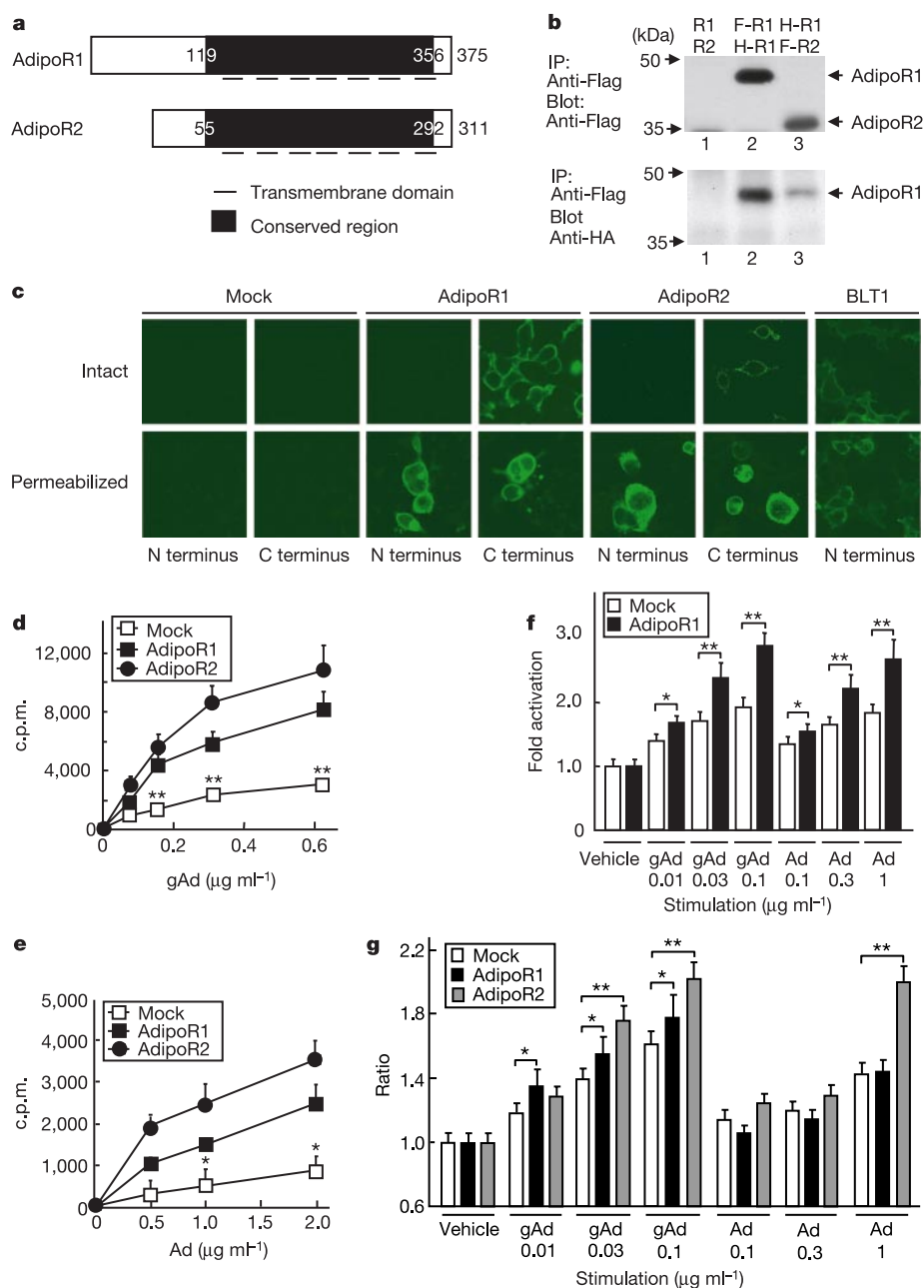
Rectangles indicate globular adiponectin binding positive cells; the cells in the rectangles were sorted in **a–c**. **d**, Northern blot analysis of AdipoR1 (top panel) and AdipoR2 (bottom panel) mRNA in mouse tissues (lanes: 1, brain; 2, heart; 3, kidney; 4, liver; 5, lung; 6, skeletal muscle; 7, spleen; 8, testis). **e**, Amino acid sequences of human and AdipoR1 and AdipoR2.

globular adiponectin (Fig. 1c), integrated cDNA of these cells was sequenced.

The cDNA analysed encoded a protein that we termed AdipoR1 (Supplementary Fig. 1c). Human and mouse AdipoR1 share 96.8% identity (Supplementary Fig. 1c). There may be two types of adiponectin receptor with distinct binding affinity for globular or full-length adiponectin<sup>10</sup> (Supplementary Fig. 1a, b). Indeed, there was one more open reading frame derived from a gene distinct from AdipoR1 in the human and mouse database<sup>15,16</sup>, which we termed AdipoR2. Human and mouse AdipoR2 share 95.2% identity

(Supplementary Fig. 1c). Human and mouse AdipoR1 is located at chromosome 1p36.13-q41 and 1 E4, whereas AdipoR2 is located at chromosome 12p13.31 and 6 F1, respectively<sup>15,16</sup>.

Northern blotting of mouse (Fig. 1d) or human tissues (Supplementary Fig. 1d) identified a major, single band of 2.0 kilobases (kb) for AdipoR1 with the predicted size<sup>15,16</sup>. AdipoR1 was expressed ubiquitously, with the most abundant expression occurring in skeletal muscle. Northern blotting analysis identified a major 4.0-kb single band for AdipoR2 with the predicted size<sup>15,16</sup>. AdipoR2 was most abundantly expressed in the mouse liver.



**Figure 2** Localization of AdipoR1 and AdipoR2, and effects of adiponectin receptor expression. **a**, Schematic structure of mouse AdipoR1 and AdipoR2. **b**, Formation of homo- or heteromultimers by human AdipoR1 and AdipoR2. F, Flag tag; H, haemagglutinin tag. **c**, Localization of AdipoR1, AdipoR2 or BLT1 (GPCR, ref. 18) with epitope tags at either end. **d–g**, Binding of <sup>125</sup>I-labelled globular (d, gAd) or full-length

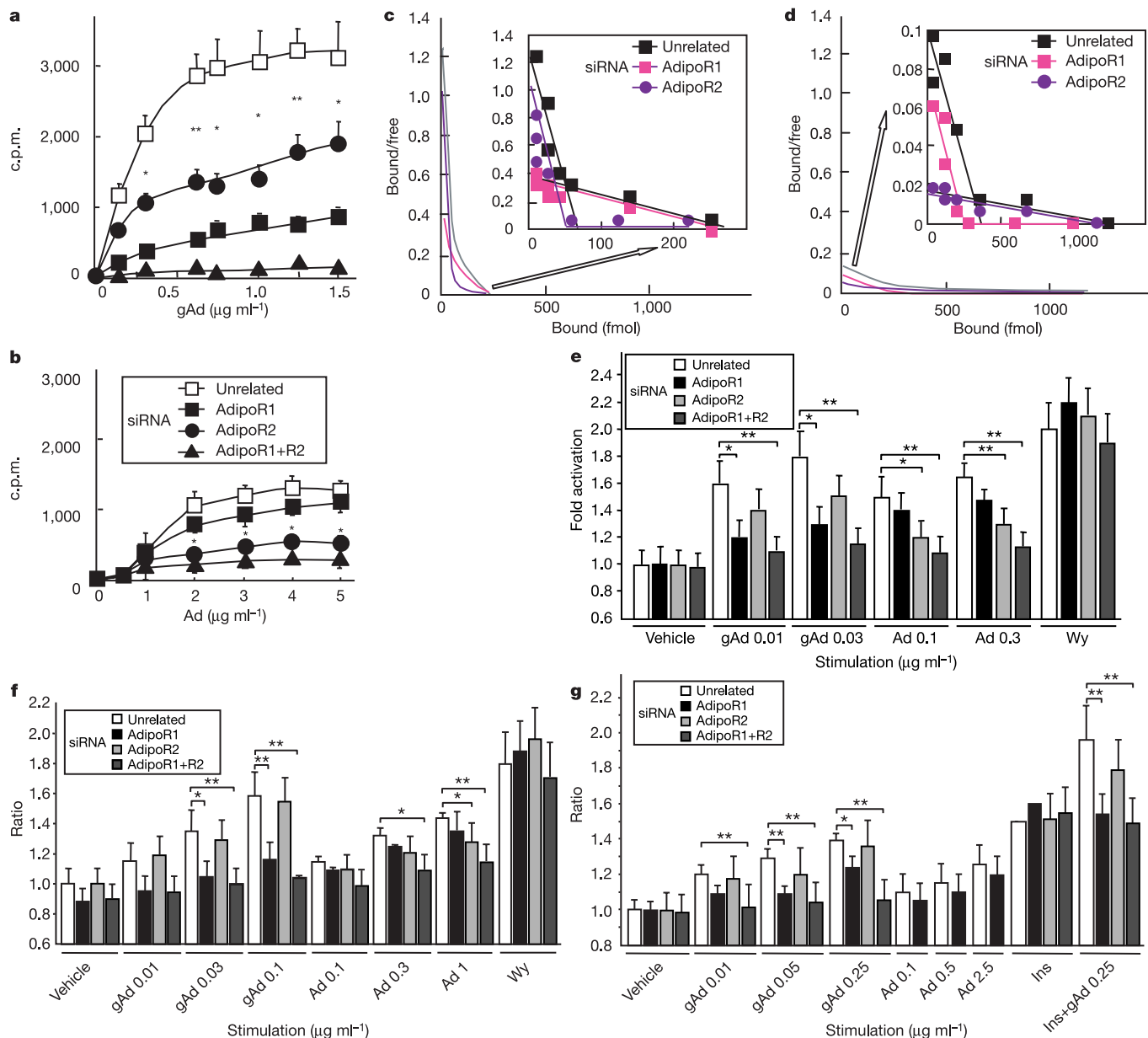
(e, Ad) adiponectin, PPAR-α ligand activity (f) and fatty-acid oxidation (g) in mouse AdipoR1- or AdipoR2-transfected C2C12 myocytes after treatment for 7 h with the indicated concentration (μg ml<sup>-1</sup>) of adiponectin. c.p.m., counts per minute. Each bar represents the mean ± s.e.m. (n = 4–6). Asterisk, P < 0.05; double asterisk, P < 0.01; between the two groups indicated, or compared with untreated.



Mouse AdipoR1 encodes a protein of 375 amino acids with the predicted molecular mass of 42.4 kDa, whereas mouse AdipoR2 encodes a protein of 311 amino acids with the predicted molecular mass of 35.4 kDa (Fig. 1e). AdipoR1 and AdipoR2 are structurally highly related—mouse AdipoR1 and AdipoR2 share 66.7% identity (Fig. 1e). AdipoR1 and AdipoR2 were predicted to encode seven transmembrane domain proteins (Fig. 2a). Although there are no other mammalian proteins that share significant sequence homology with AdipoR1 and AdipoR2 in SWISS-PROT, the proteins are conserved from yeast to human especially in the membrane-spanning regions—there are gene products in *Saccharomyces cerevisiae* (accession number Z74744), *Caenorhabditis elegans* (accession number NM\_068597) and *Drosophila* (accession number

BT001487) that show a marked homology to AdipoR1 (the human protein and these proteins share 29%, 56% and 60% identity, respectively). Notably, the yeast homologue has a principal role in metabolic pathways that regulate lipid metabolism such as fatty-acid oxidation<sup>21</sup>. Although AdipoR1 and AdipoR2 are distantly related to G-protein-coupled receptor (GPCR) families<sup>17–19</sup> in the evolutionary tree (Supplementary Fig. 1e), sequence homology of AdipoR1 and AdipoR2 to the members of GPCR families<sup>17–19</sup> is low.

Both human (Fig. 2b, top panel) and mouse (data not shown) AdipoR1 and AdipoR2 display the predicted molecular mass. We were able to detect AdipoR1 with the epitope tag haemagglutinin (HA) in anti-Flag antibody immunoprecipitates containing Flag-



**Figure 3** Effects of suppression of AdipoR1 or AdipoR2 expression by siRNA in mouse C2C12 myocytes. **a–d**, Binding (**a**, **b**) or Scatchard analysis (**c**, **d**) of <sup>125</sup>I-labelled globular (**a**, **c**) or full-length adiponectin (**b**, **d**) to C2C12 myocytes transfected with siRNA duplex. **e–g**, PPAR-α ligand activity (**e**), fatty-acid oxidation (**f**) and glucose uptake (**g**) in C2C12

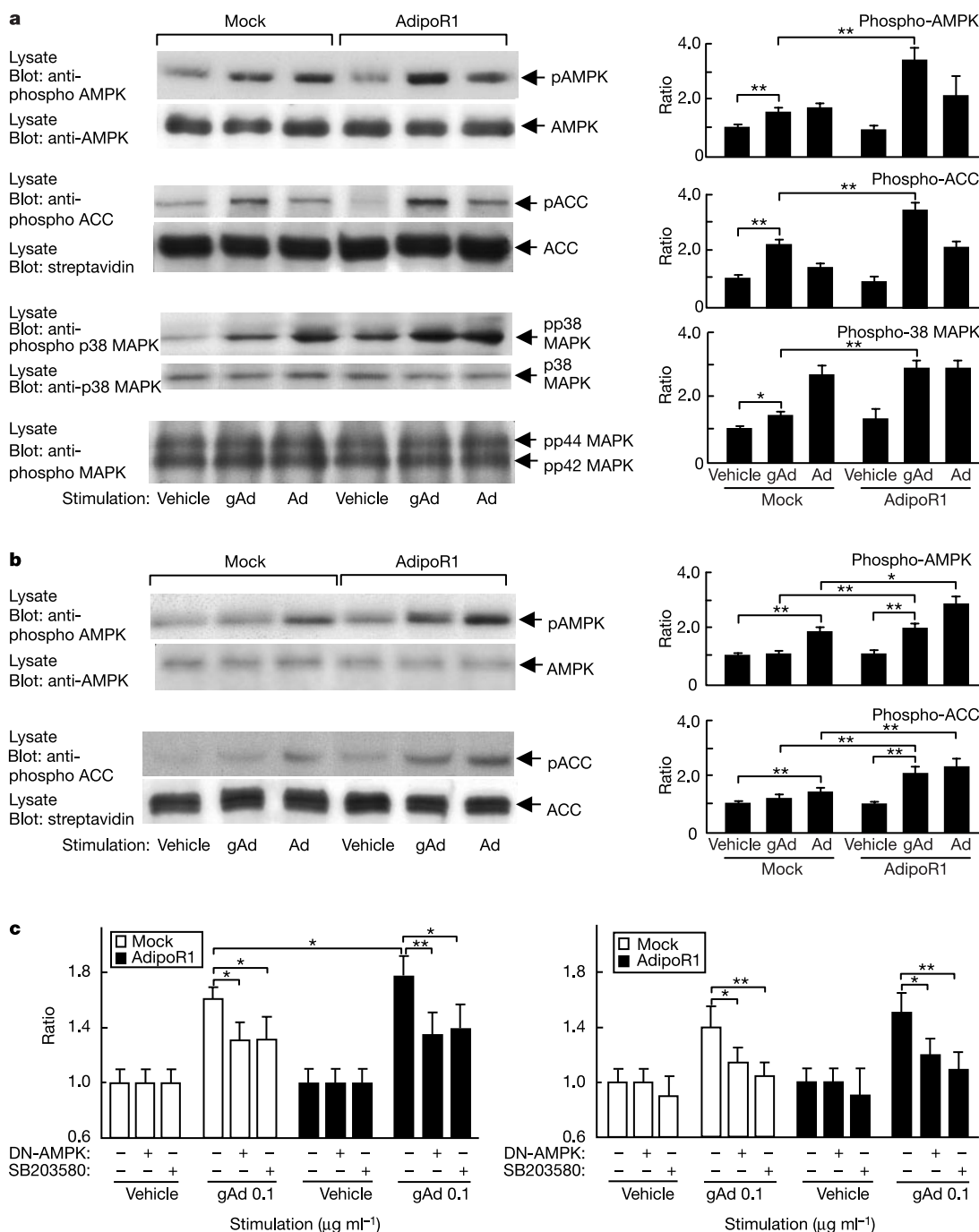
myocytes transfected with the indicated siRNA duplex. Wy, PPAR-α agonist Wy-14,643. Each bar represents the mean ± standard error ( $n = 3–8$ ). Asterisk,  $P < 0.05$ ; double asterisk,  $P < 0.01$ ; between the two groups indicated, or compared with cells transfected with unrelated siRNA.

tagged AdipoR1 and AdipoR2 (Fig. 2b, bottom panel), suggesting that AdipoR1 and AdipoR2 may be able to form both homo- and heteromultimers.

When the epitope tag was inserted at the amino terminus, we were able to detect AdipoR1 and AdipoR2 at the cell surface and intracellular organelles only when the cells were permeabilized. In contrast, when the epitope tag was inserted at the carboxy terminus we were able to detect AdipoR1 and AdipoR2 at the cell surface

without permeabilization (Fig. 2c). These data indicate that AdipoR1 and AdipoR2 are integral membrane proteins—the N terminus was internal and the C terminus was external, which is opposite to the topology of all the reported GPCR families<sup>17–19</sup>.

Expression of AdipoR1 or AdipoR2 at the cell surface in 293T cells enhanced the binding of both globular and full-length adiponectin (Supplementary Fig. 2a, b). In 293T cells expressing AdipoR1, globular or full-length adiponectin had little effect on



**Figure 4** Effects of expression of AdipoR1 on adiponectin-stimulated intracellular signals and biological effects. **a, b**, Phosphorylation and amount of AMPK (first panels), ACC (second panels), p38 MAPK (third panels) and MAPK (fourth panel) incubated with  $0.1 \mu\text{g ml}^{-1}$  globular adiponectin or  $1 \mu\text{g ml}^{-1}$  full-length adiponectin for 10 min in

C2C12 myocytes (**a**) or hepatocytes (**b**) transfected with or without AdipoR1. **c**, Fatty-acid oxidation (left panel) and glucose uptake (right panel) in mouse C2C12 myocytes. Each bar represents the mean  $\pm$  standard error ( $n = 3-5$ ). Asterisk,  $P < 0.05$ ; double asterisk,  $P < 0.01$ ; between the two groups indicated.

cyclic AMP, cyclic GMP and intracellular calcium levels (Supplementary Fig. 2c, d). In contrast, expression of AdipoR1 enhanced increases in PPAR- $\alpha$  ligand activity by both globular and full-length adiponectin (Supplementary Fig. 2e).

Expression of AdipoR1 or AdipoR2 in C2C12 myocytes (Supplementary Fig. 3a, b) enhanced both globular and full-length adiponectin binding (Fig. 2d, e), which were associated with increases in PPAR- $\alpha$  ligand activity (Fig. 2f) and fatty-acid oxidation (Fig. 2g) by both globular and full-length adiponectin. These data indicated that AdipoR1 and AdipoR2 are able to mediate the binding of globular and full-length adiponectin, as well as increases in PPAR- $\alpha$  ligand activity and fatty-acid oxidation by globular and full-length adiponectin.

C2C12 myocytes transfected with unrelated small-interfering (si)RNA<sup>20</sup> bound globular adiponectin more avidly than full-length adiponectin (Fig. 3a, b). Scatchard plot analysis revealed that there are two binding sites for globular adiponectin (Fig. 3c): high-affinity binding sites (dissociation constant ( $K_d$ ) approximately  $0.06 \mu\text{g ml}^{-1}$ , equivalent to 1.14 nM of the globular adiponectin trimer) and intermediate-affinity binding sites ( $K_d$  approximately  $0.80 \mu\text{g ml}^{-1}$ , equivalent to 14.4 nM of the globular adiponectin trimer). In contrast, there are intermediate ( $K_d$  value approximately  $6.7 \mu\text{g ml}^{-1}$ , equivalent to 49.1 nM of the full-length adiponectin hexamer) and low-affinity binding sites for full-length adiponectin ( $K_d$  value approximately  $329.3 \mu\text{g ml}^{-1}$ , equivalent to 2,415 nM of the full-length adiponectin hexamer) (Fig. 3d).

Suppression of AdipoR1 expression with siRNA (Supplementary Fig. 3c) largely reduced globular adiponectin binding (Fig. 3a) but only barely reduced full-length adiponectin binding (Fig. 3b). Scatchard plot analysis revealed that specific suppression of AdipoR1 abrogated high-affinity binding sites but failed to affect intermediate-affinity binding sites for globular adiponectin (Fig. 3c). Moreover, specific suppression of AdipoR1 only partially reduced intermediate binding sites for full-length adiponectin (Fig. 3d), and abrogated low-affinity binding sites for full-length adiponectin (Fig. 3d). In contrast to AdipoR1, suppression of AdipoR2 expression with siRNA (Supplementary Fig. 3d) largely reduced full-length adiponectin binding (Fig. 3b), but modestly reduced globular adiponectin binding (Fig. 3a). Scatchard plot analysis revealed that specific suppression of AdipoR2 barely reduced high-affinity binding sites for globular adiponectin (Fig. 3c), but abrogated intermediate-affinity binding sites (Fig. 3c). Specific suppression of AdipoR2 abrogated intermediate-affinity binding sites for full-length adiponectin (Fig. 3d), but failed to reduce low-affinity binding sites (Fig. 3d). Together, these data suggested that AdipoR1 is a high-affinity receptor for globular adiponectin and also a low-affinity receptor for full-length adiponectin, and that AdipoR2 is an intermediate-affinity receptor for full-length and globular adiponectin.

The treatment of either globular or full-length adiponectin for 7 h increased PPAR- $\alpha$  ligand activity (Fig. 3e) and stimulated fatty-acid oxidation (Fig. 3f) and glucose uptake (Fig. 3g) in C2C12 myocytes transfected with unrelated siRNA. Suppression of AdipoR1 expression by specific siRNA (Supplementary Fig. 3c) greatly reduced increases in PPAR- $\alpha$  ligand activity (Fig. 3e), fatty-acid oxidation (Fig. 3f) and glucose uptake (Fig. 3g) by globular adiponectin. In contrast, the suppression of AdipoR1 expression failed to significantly reduce these effects by full-length adiponectin (Fig. 3e–g). Suppression of AdipoR2 expression by specific siRNA (Supplementary Fig. 3d) partially reduced increases in PPAR- $\alpha$  ligand activity (Fig. 3e) and fatty-acid oxidation (Fig. 3f) by full-length adiponectin, but not by globular adiponectin.

Suppression of both AdipoR1 and AdipoR2 expression in combination by siRNA in C2C12 myocytes (Supplementary Fig. 3c, d) almost abolished globular adiponectin binding (Fig. 3a) and largely abolished full-length adiponectin binding (Fig. 3b), and at the same time almost abolished the increased PPAR- $\alpha$  ligand activity

(Fig. 3e), fatty-acid oxidation (Fig. 3f) and glucose uptake (Fig. 3g) by globular or full-length adiponectin.

Mock-transfected hepatocytes specifically bound full-length adiponectin (Supplementary Fig. 3f) but not globular adiponectin (Supplementary Fig. 3e). Expression of AdipoR1 or AdipoR2 alone or in combination made it possible for globular adiponectin to bind to hepatocytes (Supplementary Fig. 3e) and further increased full-length adiponectin binding to hepatocytes (Supplementary Fig. 3f), both of which were associated with increases in PPAR- $\alpha$  ligand activity by globular and full-length adiponectin (Supplementary Fig. 3g). Conversely, suppression of AdipoR2 expression by siRNA in hepatocytes significantly reduced full-length adiponectin binding (Supplementary Fig. 3h) as well as increases in PPAR- $\alpha$  ligand activity by full-length adiponectin (Supplementary Fig. 3i).

In mock-transfected C2C12 myocytes both globular and full-length adiponectin increased the amount of phosphorylation of AMP kinase (AMPK), acetyl coenzyme A carboxylase (ACC)<sup>10,11</sup> and p38 mitogen-activated protein kinase (MAPK)<sup>22–24</sup>, but not other protein kinases such as MAPK (Fig. 4a). Expression of AdipoR1 in C2C12 myocytes was associated with increased phosphorylation of AMPK, ACC and p38 MAPK on stimulation with globular adiponectin (Fig. 4a), suggesting that AdipoR1 can mediate globular adiponectin-stimulated AMPK<sup>10,11</sup> and p38 MAPK activation<sup>22–24</sup> (Fig. 4a).

In mock-transfected hepatocytes, full-length but not globular adiponectin stimulated AMPK activation and ACC phosphorylation (Fig. 4b). Expression of AdipoR1 in hepatocytes was associated with increases in AMPK and ACC phosphorylation on stimulation with both globular and full-length adiponectin (Fig. 4b), suggesting that AdipoR1 can mediate globular and full-length adiponectin-stimulated AMPK and ACC phosphorylation.

In mock-transfected C2C12 myocytes, globular adiponectin-stimulated fatty-acid oxidation and glucose uptake were partially inhibited by dominant negative (DN)-AMPK or SB203580, a specific inhibitor of p38 MAPK<sup>22–24</sup> (Fig. 4c). Expression of AdipoR1 in C2C12 myocytes further increased fatty-acid oxidation and glucose uptake on stimulation with globular adiponectin; these effects were also inhibited partially by DN-AMPK or SB203580 (Fig. 4c). Thus, the stimulation of fatty-acid oxidation and glucose uptake by globular adiponectin through AdipoR1 seemed to be dependent on both AMPK and p38 MAPK pathways in C2C12 myocytes.

In this study we have isolated cDNA encoding the adiponectin receptors AdipoR1 and AdipoR2. Expression of these receptors or their suppression supports our conclusion that they serve as receptors for globular and full-length adiponectin, and that they mediate increased AMPK, PPAR- $\alpha$  ligand activity, the fatty-acid oxidation and glucose uptake by adiponectin.

Scatchard plot analyses showed that AdipoR1 is a high-affinity receptor for globular adiponectin but a very low-affinity receptor for full-length adiponectin, and that AdipoR2 is an intermediate affinity receptor for globular and full-length adiponectin. In this respect, fatty-acid oxidation mediated through AdipoR1 was highly sensitive to globular adiponectin (half-maximal effective dose ( $ED_{50}$ ) approximately  $0.03 \mu\text{g ml}^{-1}$ , equivalent to 0.54 nM of the globular adiponectin trimer) but was resistant to full-length adiponectin. Fatty-acid oxidation mediated through AdipoR2 was intermediately sensitive to globular or full-length adiponectin ( $ED_{50}$  approximately  $0.85 \mu\text{g ml}^{-1}$ , equivalent to 6.23 nM of the full-length adiponectin hexamer). Thus, there was a good correlation between binding affinity (Fig. 3c, d) and adiponectin sensitivity (Fig. 2g), and the  $ED_{50}$  corresponded to 13–50% of the  $K_d$  on a molar basis.

Interactions between adiponectin and AdipoR1 increased PPAR- $\alpha$ , AMPK and p38 MAPK activation. PPAR- $\alpha$  activation is supposedly involved in adiponectin-stimulated fatty-acid oxidation, but



not glucose uptake. AMPK or p38 MAPK activation may be involved in adiponectin-stimulated fatty-acid oxidation and glucose uptake, as both DN-AMPK and SB203580 partially inhibited adiponectin-stimulated fatty-acid oxidation and glucose uptake in an AdipoR1 expression-dependent fashion. p38 MAPK has been reported to activate PPAR- $\alpha$  through increased phosphorylation of PPAR- $\alpha$ <sup>22</sup> and increased co-activation specifically by PGC-1 (refs 22, 23), thereby stimulating fatty-acid oxidation. p38 MAPK has also been reported to stimulate glucose transport by means of increased MEF2 transcriptional activity through an increase of PGC-1 phosphorylation and activation<sup>24</sup>, but not through PPAR- $\alpha$ . However, because the extent of AMPK or p38 MAPK activation by globular or full-length adiponectin cannot fully explain the extent of their biological effects, other signalling pathways also seem to be involved. Finally, identification of the 'missing link' between adiponectin receptors and adiponectin-activated protein kinases is an important next step towards our understanding of the actions of adiponectin.

AdipoR1 and AdipoR2 are predicted to contain seven transmembrane domains, but are structurally, topologically and functionally distinct from GPCRs. They do not seem to be coupled with G protein, but activate unique sets of signalling molecules such as PPAR- $\alpha$ , AMPK and p38 MAPK. Thus, adiponectin receptors may comprise a new receptor family. AdipoR1 and AdipoR2 encode receptors for globular and full-length adiponectin that mediate antidiabetic metabolic effects. Molecular cloning of AdipoR1 and AdipoR2 should facilitate the understanding of molecular mechanisms of adiponectin actions and obesity-linked diseases, such as diabetes and atherosclerosis, and the designing of new antidiabetic and anti-atherogenic drugs with AdipoR1 and AdipoR2 as molecular targets. □

## Methods

### FACS analysis and sequencing of integrated retroviruses

10<sup>7</sup> Plat-E packaging cells<sup>25</sup> were transiently transfected with 10  $\mu$ g human skeletal muscle cDNA library (Clontech) using Lipofectamine Plus (Life Technologies). Ba/F3 cells were infected with 1/20-diluted supernatants corresponding to an estimated multiplicity of infection of 0.3 (ref. 14). For selection, we performed fluorescence-activated cell sorting (FACS) analysis<sup>14</sup>.

### Quantitative analysis of AdipoR1/R2 transcripts

For quantification of adiponectin receptor mRNAs we used the real-time polymerase chain reaction method and northern blot<sup>10,18</sup> using cDNA probes corresponding to the *Pst*I–*Bst*XI, *Bam*HI–*Pst*II, or *Eco*RV–*Not*I fragment of human AdipoR1, mouse AdipoR1 or human and mouse AdipoR2 cDNA, respectively. The primer sets and the probes for mouse AdipoR1/R2 and human AdipoR1/R2 were as follows: mouse AdipoR1 forward primer ACGTTGGAGAGTCATCCCGTAT, reverse primer CTCTGTGTGGATGCGGAAGAT and the probe CCTGCTACATGGCCACAGACCACTT with a minor groove binder; mouse AdipoR2 forward primer TCCCAGGAAGATGAAGGGTTTAT, reverse primer TTCCATTCGTTTCATAGCATGA and the probe ATGTCCCGCTCTACAGGCC with a minor groove binder; human AdipoR1 forward primer TTCTTCCTCATGGCTGTGATGT, reverse primer AAGAAGCGCTCAGGAATTCG and the probe TCACTGGAGCTGGCCTTTATGCTGC with a minor groove binder; human AdipoR2 forward primer ATAGGGCAGATAGGCTGGTTGA, reverse primer GGATCCGGGCAGCATACA and the probe CTGATGGCCAGCCTCTACATCACAGGA with a minor groove binder.

### Recombinant and purification of circulating adiponectin

Bacterially expressed murine full-length adiponectin and globular adiponectin were generated<sup>10</sup>, and circulating adiponectin was purified<sup>47</sup>. Immunoblot analysis after cross-linking using BS<sup>3</sup> indicated that recombinant full-length or circulating adiponectin forms monomers, trimers, hexamers and species of high molecular mass (Supplementary Fig. 4g), and that recombinant globular adiponectin exists as both a monomer and trimer<sup>10</sup>. No significant differences were observed between recombinant full-length adiponectin and circulating adiponectin in the binding to AdipoR1 and AdipoR2 in C2C12 myocytes (Fig. 3b; see also Supplementary Fig. 4i) and in the biological activity such as stimulation of p38 MAPK phosphorylation (Supplementary Fig. 4h).

### Expression in mammalian cells and characterization

The AdipoR1 or AdipoR2 expression vector was constructed by ligating into the *Eco*RV–*Not*I site of pCXN2 (ref. 26). DNA transfection was performed by lipofection using Lipofectamine Plus (Life Technologies). The cellular location of AdipoR1 or AdipoR2 was evaluated by confocal fluorescence microscopy using 293T cells. Intracellular Ca<sup>2+</sup> concentration and cAMP and cGMP contents<sup>18</sup>, phosphorylation and amount of AMPK,

ACC<sup>10</sup>, p38 MAPK and MAPK<sup>22–24</sup>, PPAR- $\alpha$  ligand activity<sup>12</sup>, [<sup>14</sup>C]CO<sub>2</sub> production from [<sup>1-14</sup>C]palmitic acid, and glucose uptake<sup>6,10</sup> were determined.

### Binding assay

Recombinant globular or full-length adiponectin was biotinylated with NHS-LC-biotin (Pierce) (Supplementary Fig. 1a, b). Synthetic adiponectin was labelled with [<sup>125</sup>I] at Tyr by IODO beads (Pierce) in the presence of Na[<sup>125</sup>I] (2,000 Ci mmol<sup>–1</sup>, Amersham Pharmacia Biotech) according to the manufacturer's protocol (Figs 2 and 3; see also Supplementary Figs 2–4). Cells were seeded at a density of 4.1  $\times$  10<sup>4</sup> cells per well. After an overnight culture, the cells were incubated at 4 °C for 1 h with binding buffer (ice-cold phosphate-buffered saline (PBS), 0.1% bovine serum albumin) containing designated concentrations of [<sup>125</sup>I]-labelled adiponectin (5,000 counts per min per ng protein) plus unlabelled competitors. The binding equilibrium was found to be established when the binding assay was conducted at 4 °C after 1 h. The cells were then washed three times with ice-cold PBS, lysed in 0.1 M NaOH, 0.1% SDS, and the cell-bound radioactivity was determined using a  $\gamma$ -counter<sup>10,18</sup>. Nonspecific binding was determined using a 200-fold excess of unlabelled adiponectin. Specific binding was calculated by subtracting nonspecific binding from the total binding. The values presented in this work represent an average of triplicate determinations of 3–10 experiments. Because of high homology between adiponectin and C1q, we studied C1q binding and found that C1q did not bind to AdipoR1 or AdipoR2 in C2C12 myocytes (data not shown).

### RNA interference

Two pairs of siRNAs were chemically synthesized, annealed and transfected into 60–70% confluent C2C12 myocytes, hepatocytes or human aortic endothelial cells using Lipofectamine Plus (Life Technologies)<sup>20</sup>. The sequences of the sense siRNAs are as follows: mouse AdipoR1, GAGACUGGCAACAUCUGGACATT; mouse AdipoR2, GCUUAGAGACACCUUUUUUUTT; human AdipoR1, GGACAACGACUAUCUGUACATT; human AdipoR2, GGAGUUUCGUUUAUGAUCGGTT. Forty-eight hours after transfection, the cells were lysed.

### Predicted structure of AdipoR1 and AdipoR2

Hydropathy plots of the adiponectin receptor proteins were conducted using the hydrophobicity indices of Kyte and Doolittle<sup>27</sup>. We examined whether AdipoR1 and AdipoR2 have homology to any other class of GPCR by the method described at <http://cbrg.inf.ethz.ch/Server/AllAll.html>.

Received 31 December 2002; accepted 1 May 2003; doi:10.1038/nature01705.

- Scherer, P. E., Williams, S., Fogliano, M., Baldini, G. & Lodish, H. F. A novel serum protein similar to C1q, produced exclusively in adipocytes. *J. Biol. Chem.* **270**, 26746–26749 (1995).
- Hu, E., Liang, P. & Spiegelman, B. M. AdipoQ is a novel adipose-specific gene dysregulated in obesity. *J. Biol. Chem.* **271**, 10697–10703 (1996).
- Maeda, K. *et al.* cDNA cloning and expression of a novel adipose specific collagen-like factor, apM1 (AdiPose Most abundant Gene transcript 1). *Biochem. Biophys. Res. Commun.* **221**, 286–296 (1996).
- Nakano, Y., Tobe, T., Choi-Miura, N. H., Mazda, T. & Tomita, M. Isolation and characterization of GBP28, a novel gelatin-binding protein purified from human plasma. *J. Biochem. (Tokyo)* **120**, 802–812 (1996).
- Fruebis, J. *et al.* Proteolytic cleavage product of 30-kDa adipocyte complement-related protein increases fatty acid oxidation in muscle and causes weight loss in mice. *Proc. Natl Acad. Sci. USA* **98**, 2005–2010 (2001).
- Yamauchi, T. *et al.* The fat-derived hormone adiponectin reverses insulin resistance associated with both lipodystrophy and obesity. *Nature Med.* **7**, 941–946 (2001).
- Berg, A. H., Combs, T. P., Du, X., Brownlee, M. & Scherer, P. E. The adipocyte-secreted protein Acrp30 enhances hepatic insulin action. *Nature Med.* **7**, 947–953 (2001).
- Kubota, N. *et al.* Disruption of adiponectin causes insulin resistance and neointimal formation. *J. Biol. Chem.* **277**, 25863–25866 (2002).
- Maeda, N. *et al.* Diet-induced insulin resistance in mice lacking adiponectin/ACRP30. *Nature Med.* **8**, 731–737 (2002).
- Yamauchi, T. *et al.* Adiponectin stimulates glucose utilization and fatty-acid oxidation by activating AMP-activated protein kinase. *Nature Med.* **8**, 1288–1295 (2002).
- Tomas, E. *et al.* Enhanced muscle fat oxidation and glucose transport by ACRP30 globular domain: acetyl-CoA carboxylase inhibition and AMP-activated protein kinase activation. *Proc. Natl Acad. Sci. USA* **99**, 16309–16313 (2002).
- Yamauchi, T. *et al.* Globular adiponectin protected ob/ob mice from diabetes and apoE deficient mice from atherosclerosis. *J. Biol. Chem.* **278**, 2461–2468 (2003).
- Ouchi, N. *et al.* Adipocyte-derived plasma protein, adiponectin, suppresses lipid accumulation and class A scavenger receptor expression in human monocyte-derived macrophages. *Circulation* **103**, 1057–1063 (2001).
- Kitamura, T. *et al.* Efficient screening of retroviral cDNA expression libraries. *Proc. Natl Acad. Sci. USA* **92**, 9146–9150 (1995).
- Waterston, R. H. *et al.* Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562 (2002).
- Okazaki, Y. *et al.* Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature* **420**, 563–573 (2002).
- Wess, J. G-protein-coupled receptors: molecular mechanisms involved in receptor activation and selectivity of G-protein recognition. *FASEB J.* **11**, 346–354 (1997).
- Yokomizo, T., Izumi, T., Chang, K., Takuwa, Y. & Shimizu, T. A G-protein-coupled receptor for leukotriene B<sub>4</sub> that mediates chemotaxis. *Nature* **387**, 620–624 (1997).
- Scheer, A., Fanelli, F., Costa, T., De Benedetti, P. G. & Cotechia, S. Constitutively active mutants of the alpha 1B-adrenergic receptor: role of highly conserved polar amino acids in receptor activation. *EMBO J.* **15**, 3566–3578 (1996).

20. Miyagishi, M. & Taira, K. U6 promoter-driven siRNAs with four uridine 3' overhangs efficiently suppress targeted gene expression in mammalian cells. *Nature Biotechnol.* **20**, 497–500 (2002).
21. Karpichev, I. V., Cornivelli, L. & Small, G. M. Multiple regulatory roles of a novel *Saccharomyces cerevisiae* protein, encoded by YOL002c, in lipid and phosphate metabolism. *J. Biol. Chem.* **277**, 19609–19617 (2002).
22. Barger, P. M., Browning, A. C., Garner, A. N. & Kelly, D. P. p38 mitogen-activated protein kinase activates peroxisome proliferator-activated receptor  $\alpha$ : a potential role in the cardiac metabolic stress response. *J. Biol. Chem.* **276**, 44495–44501 (2001).
23. Puigserver, P. *et al.* Cytokine stimulation of energy expenditure through p38 MAP kinase activation of PPAR $\gamma$  coactivator-1. *Mol. Cell* **8**, 971–982 (2001).
24. Michael, L. F. *et al.* Restoration of insulin-sensitive glucose transporter (GLUT4) gene expression in muscle cells by the transcriptional coactivator PGC-1. *Proc. Natl Acad. Sci. USA* **98**, 3820–3825 (2001).
25. Morita, S., Kojima, T. & Kitamura, T. Plat-E: an efficient and stable system for transient packaging of retroviruses. *Gene Ther.* **7**, 1063–1066 (2000).
26. Niwa, H., Yamamura, K. & Miyazaki, J. Efficient selection for high-expression transfectants with a novel eukaryotic vector. *Gene* **108**, 193–200 (1991).
27. Kyte, J. & Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132 (1982).

**Supplementary Information** accompanies the paper on [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We are grateful to K. Kirii, A. Itoh, A. Okano, T. Nagano and S. Nakamura for their technical assistance. This work was supported by a Grant-in-Aid for Creative Scientific Research from the Japan Society for the Promotion of Science (to T.K.), and by Health Science Research Grants (Research on Human Genome and Gene Therapy) from the Ministry of Health and Welfare (to T.K.).

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to T.K. (kadowaki-3im@h.u-tokyo.ac.jp). The GenBank accession numbers for human and mouse AdipoR1 are NM\_015999 and BCO14875, and for human and mouse AdipoR2 are NM\_024551 and XM\_132831, respectively.

## Redox regulation of protein tyrosine phosphatase 1B involves a sulphenyl-amide intermediate

Annette Salmeen<sup>\*†</sup>, Jannik N. Andersen<sup>‡</sup>, Michael P. Myers<sup>‡</sup>, Tzu-Ching Meng<sup>‡</sup>, John A. Hinks<sup>\*</sup>, Nicholas K. Tonks<sup>‡</sup> & David Barford<sup>\*</sup>

<sup>\*</sup> Section of Structural Biology, Institute of Cancer Research, Chester Beatty Laboratories, 237 Fulham Road, London SW3 6JB, UK

<sup>‡</sup> Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA

<sup>†</sup> Present address: Department of Molecular Pharmacology, Stanford University Medical School, 269 Campus Drive, Stanford University, Stanford, California 94305, USA

The second messenger hydrogen peroxide is required for optimal activation of numerous signal transduction pathways, particularly those mediated by protein tyrosine kinases<sup>1–6</sup>. One mechanism by which hydrogen peroxide regulates cellular processes is the transient inhibition of protein tyrosine phosphatases through the reversible oxidization of their catalytic cysteine, which suppresses protein dephosphorylation<sup>7–9</sup>. Here we describe a structural analysis of the redox-dependent regulation of protein tyrosine phosphatase 1B (PTP1B), which is reversibly inhibited by oxidation after cells are stimulated with insulin<sup>8</sup> and epidermal growth factor<sup>9</sup>. The sulphenic acid intermediate produced in response to PTP1B oxidation is rapidly converted into a previously unknown sulphenyl-amide species, in which the sulphur atom of the catalytic cysteine is covalently linked to the main chain nitrogen of an adjacent residue. Oxidation of PTP1B to the sulphenyl-amide form is accompanied by large

conformational changes in the catalytic site that inhibit substrate binding. We propose that this unusual protein modification both protects the active-site cysteine residue of PTP1B from irreversible oxidation to sulphinic acid and permits redox regulation of the enzyme by promoting its reversible reduction by thiols.

Protein tyrosine phosphatases (PTPs) are cysteine-dependent enzymes defined by a PTP signature motif, Cys-(Xaa)<sub>5</sub>-Arg, at their catalytic site<sup>10</sup>. Owing to its unique environment, the catalytic cysteine has a low pK<sub>a</sub>, which enhances its function as a nucleophile and renders PTPs susceptible to inactivation by reactive oxygen species such as hydrogen peroxide (H<sub>2</sub>O<sub>2</sub>) and superoxide<sup>11</sup>. To understand the mechanism of the reversibility of PTP1B redox regulation, we integrated X-ray crystallography and mass spectrometry to monitor the changes in the enzyme in response to oxidation. We sought to determine the structure of PTP1B with the catalytic cysteine in the physiologically relevant, reversibly oxidized Cys-SOH (sulphenic acid) state.

After incubating the enzyme with stoichiometric amounts of H<sub>2</sub>O<sub>2</sub>, however, we did not observe formation of sulphenic acid but instead identified an unexpected modification of the catalytic site cysteine residue, which we call a 'sulphenyl-amide' species (Fig. 1a). In addition, when we determined the structure of PTP1B at several stages during a time course of oxidation in response to H<sub>2</sub>O<sub>2</sub>, only the newly identified sulphenyl-amide intermediate was detected, and not the Cys-SOH state (Fig. 1b).

In this time course, almost complete conversion of PTP1B to the sulphenyl-amide species occurred by 2 h, and a similar structure was observed at 5 h. At earlier time points, the electron density maps were consistent with a mixture of both reduced and sulphenyl-amide conformational states. Notably, there was no definitive evidence of a sulphenic acid structure, indicating that the rate of formation of this intermediate is rate limiting in the generation of the sulphenyl-amide species. Alternative conformation refinement with REFMAC<sup>12</sup> accounted for the observed electron density maps, allowing the relative proportions of the reduced and sulphenyl-amide states to be estimated (Supplementary Table 1). After a 16-h incubation, the sulphenyl-amide species was oxidized to a mixture of sulphinic (Cys-SO<sub>2</sub>H) and sulphonic (Cys-SO<sub>3</sub>H) acids, and this structure of PTP1B resembled PTP1B-SO<sub>3</sub>H generated with pervanadate, a strong oxidizing agent (Supplementary Information).

The oxidation of the catalytic cysteine to the sulphenyl-amide state was accompanied by profound changes in the structure of the active site and inhibition of activity. In reduced PTP1B, the catalytic Cys 215 is at the base of a cleft on the surface of the protein formed by residues of the PTP signature motif (residues 214–222), which create the PTP loop (refs 10, 13 and Fig. 2a). The thiol group of Cys 215 is poised for nucleophilic attack onto the substrate, and the S $\gamma$  atom of Cys 215 accepts a hydrogen bond from the side chain of Ser 222. In the sulphenyl-amide state, electron density maps in the region of the PTP loop indicate well-defined continuous density bridging the S $\gamma$ -atom of Cys 215 and the main chain nitrogen atom of Ser 216 (Figs 1 and 2b). The sulphenyl-amide species is probably formed by nucleophilic attack of the main chain nitrogen atom of Ser 216 onto the electrophilic sulphur atom of the labile sulphenic acid, with concomitant release of water (Fig. 2c). Notably,  $F_o - F_c$  electron density maps show that there are no additional oxygen atoms attached to S $\gamma$  of Cys 215 in the sulphenyl-amide species (Fig. 1). This is consistent with the occurrence of a nucleophilic substitution reaction (Fig. 2c).

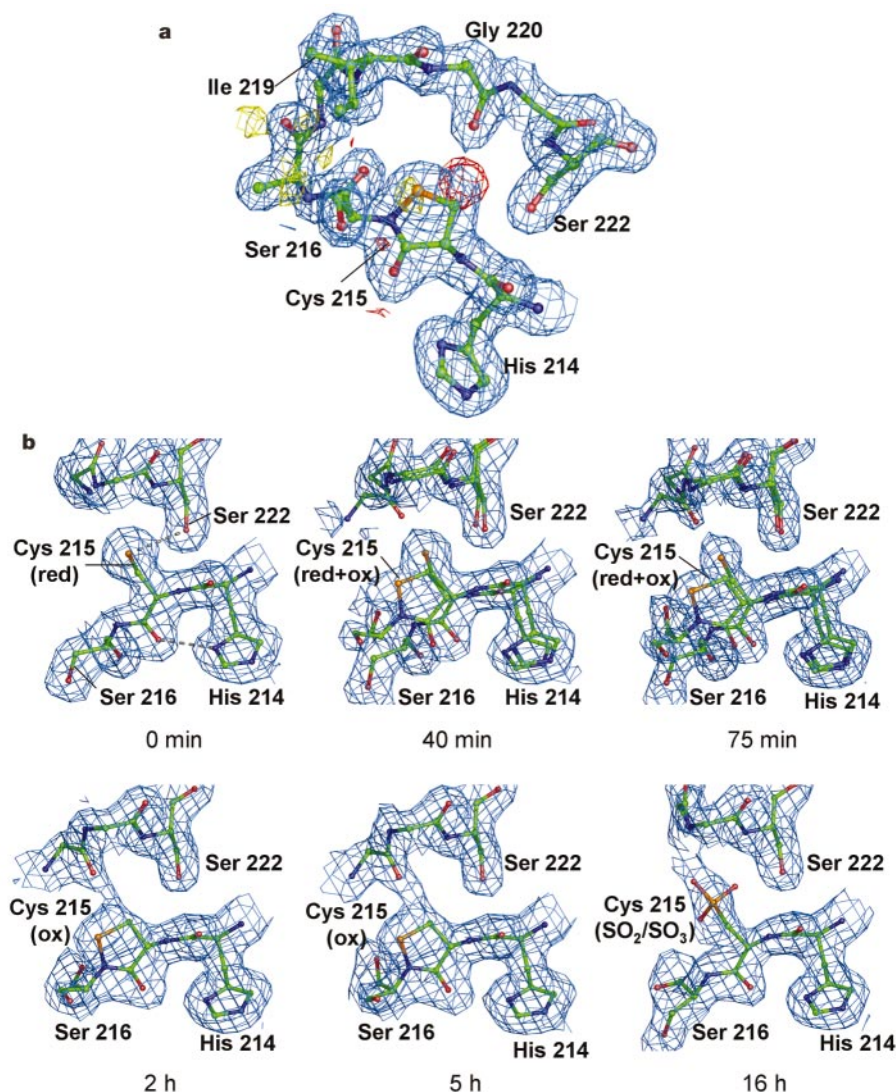
Formation of the sulphenyl-amide bond, which imposes conformational constraints on the main chain of the PTP loop and disruption of the Cys 215–Ser 222 hydrogen bond, triggers tertiary structural changes in the protein (Fig. 2d). Notably, the PTP loop undergoes a marked change in conformation. Gly 218 shifts by about 7 Å, and the helical conformation of the PTP loop in reduced PTP1B converts to a reverse  $\beta$ -hairpin conformation. The phosphotyrosine (pTyr) loop containing Tyr 46, which defines the depth of

the catalytic cleft<sup>14</sup>, also undergoes a significant conformational change. In reduced PTP1B, Tyr 46 is directed towards the PTP loop and its phenolic hydroxyl group forms a hydrogen bond with the side chain of Ser 216 of the PTP loop (Fig. 2a). The conformational shift of the Ser 216 side chain in the sulphenyl-amide structure disrupts the hydrogen bond to the hydroxyl group of Tyr 46 and promotes a displacement of Tyr 46 to a solvent-exposed position, accompanied by shifts of the whole pTyr recognition loop.

In the Q loop (a third loop surrounding the PTP catalytic site), the side chain of Gln 262, which mediates hydrolysis of the cysteinyl phosphate catalytic intermediate<sup>15</sup>, changes orientation. The WPD loop, containing the essential Asp 181 residue that functions in both stages of catalysis, remains unchanged, as does the rest of the protein. The open conformation of the PTP loop in the sulphenyl-amide state exposes the S $\gamma$  atom of Cys 215, enhancing its accessibility for modification by thiols (Fig. 2b). Notably, incubating crystals of sulphenyl-amide PTP1B with either dithiothreitol (DTT) or glutathione (GSH) led to quantitative reduction of the

sulphenyl-amide cysteine and to a conformational change in both the PTP and pTyr loops to their catalytically active conformations (Supplementary Information). Further oxidation of the sulphenyl-amide bond to sulphonic acid also reversed the changes in tertiary structure (data not shown), indicating that the open conformations of the PTP and pTyr loops are associated only with the sulphenyl-amide species of the catalytic cysteine.

To explore the oxidation of PTP1B in solution, we subjected the protein to varying amounts of H<sub>2</sub>O<sub>2</sub> and then, after trypsinization, examined the mass of the peptide containing the active site cysteine by matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry (Fig. 3a). At high concentrations of H<sub>2</sub>O<sub>2</sub>, PTP1B was irreversibly oxidized and a shift in relative molecular mass ( $M_r$ ) of +32 or +48 (corresponding to the formation of sulphinic or sulphonic acid derivatives, respectively) was readily detected. At lower concentrations of H<sub>2</sub>O<sub>2</sub>, which reversibly oxidized PTP1B, we did not detect the formation of a Cys-SOH modification. Similar results were found for the intact



**Figure 1** Oxidation of PTP1B results in formation of a sulphenyl-amide bond between Cys 215 and Ser 216. **a**, The PTP loop in the sulphenyl-amide structure. Cyan, red and yellow correspond to  $2F_o - F_c$  and positive and negative  $F_o - F_c$  electron density maps, respectively. The  $F_o - F_c$  density indicates that there are no oxygen atoms attached to the Cys 215 S $\gamma$  atom, although a small amount of the enzyme is still reduced. **b**, Time

course of PTP1B oxidation. Electron density maps ( $2F_o - F_c$ ) show the time-dependent changes at the catalytic Cys 215 of PTP1B over a 16-h period. At 40 and 75 min, there is a mixture of reduced and oxidized states. H<sub>2</sub>O<sub>2</sub> was in 100- to 1,000-fold molar excess over PTP1B. red, reduced; ox, sulphenyl-amide structures. All figures were drawn with PYMOL (see <http://pymol.sourceforge.net/>).

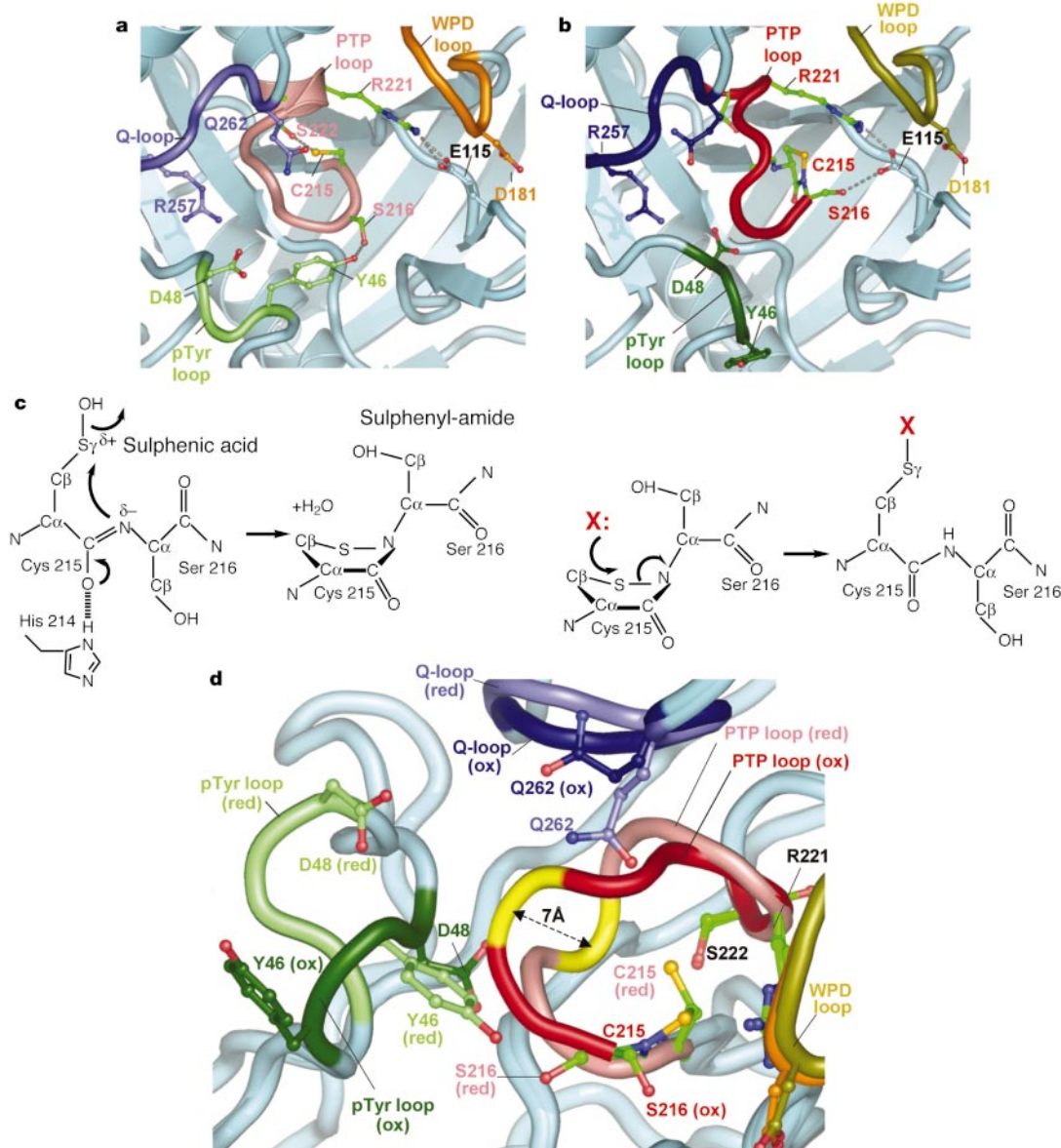


protein (Supplementary Information). In the PTP1B C215S mutant, in which the active-site cysteine is replaced by serine, there was no change in the mass of this peptide at any concentration of  $H_2O_2$  (Fig. 3a).

To test whether the failure to detect the sulphenic acid derivative in the wild-type protein was due to its rapid conversion to the sulphenyl-amide species, we carried out a pulse-chase analysis with  $H_2^{18}O_2$ . As expected, incubation of PTP1B with a 1,000-fold molar excess 'heavy'  $H_2^{18}O_2$  or 'light'  $H_2^{16}O_2$  (50 mM) irreversibly inactivated PTP1B (Fig. 3b, left) and led to the production of sulphenic acid derivatives with a mass increase of +54 (all  $^{18}O$ ) or +48 (all  $^{16}O$ ), respectively (Fig. 3b, right). However, when PTP1B was pre-incubated with concentrations of  $H_2^{18}O_2$  that reversibly oxidized PTP1B and then chased with a more than 100-fold excess of  $H_2^{16}O_2$ , we did not detect production of a sulphenic acid derivative of mass +50 (one  $^{18}O$  and two  $^{16}O$ ) but only +48 (all  $^{16}O$ ) (Fig. 3b). This observation is consistent with the notion that at concentrations of

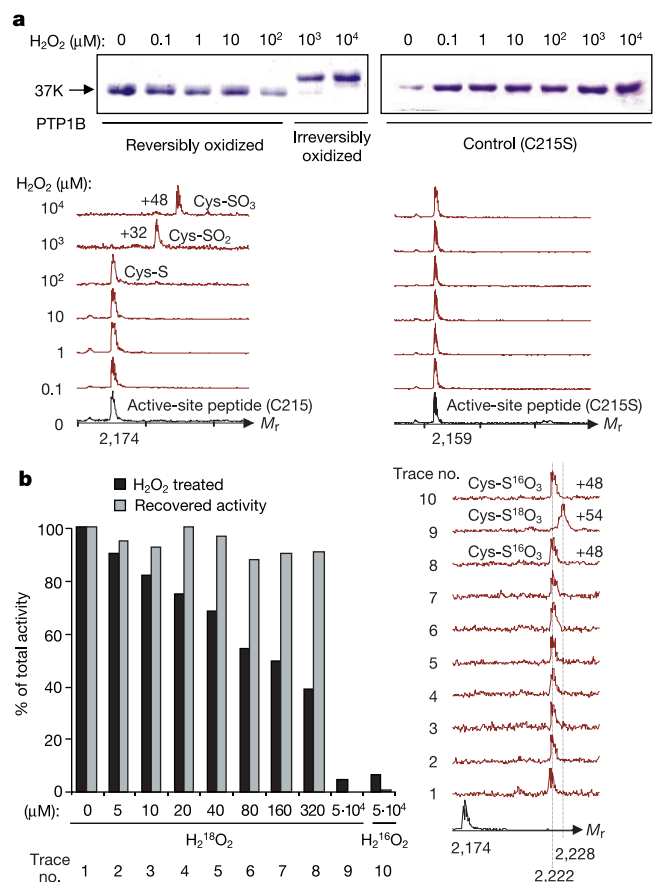
peroxide that result in reversible inhibition of PTP1B, oxygen is rapidly eliminated from sulphenic acid—the first oxidized form of the active-site cysteine—to generate a more stable derivative, such as the sulphenyl-amide species (Fig. 2c). Notably, this species is formed in the presence of thiols, suggesting that it would form in the reducing environment of a cell.

Together, our crystallographic and pulse-chase mass spectrometry data establish a model for the interrelationship of PTP1B oxidation states (Fig. 4a). Oxidation to the sulphenyl-amide state would be anticipated to inhibit PTP activity in a reversible manner, with the chemical and structural characteristics of the sulphenyl-amide species both suppressing irreversible oxidation to sulphinic and sulphonic acids and permitting modification of Cys 215 by thiols to regenerate an active state of the enzyme. To examine the effects of the oxidation-mediated structural changes on the function of PTP1B, we incubated a substrate trapping mutant of PTP1B (D181A)<sup>16</sup> with the kinase domain of



**Figure 2** Conformational changes accompanying the oxidation of PTP1B. **a**, Ribbon diagram showing catalytic site of reduced PTP1B. The PTP loop is shown in red. **b**, Sulphenyl-amide species of PTP1B in the same orientation as in **a**. **c**, Chemical

mechanism for generating the sulphenyl-amide bond. 'X' denotes a nucleophile. **d**, Superimposition of reduced (red) and sulphenyl-amide (ox) states of PTP1B. Gly 218 is shown in yellow. For clarity, the view has been rotated relative to **a** and **b**.

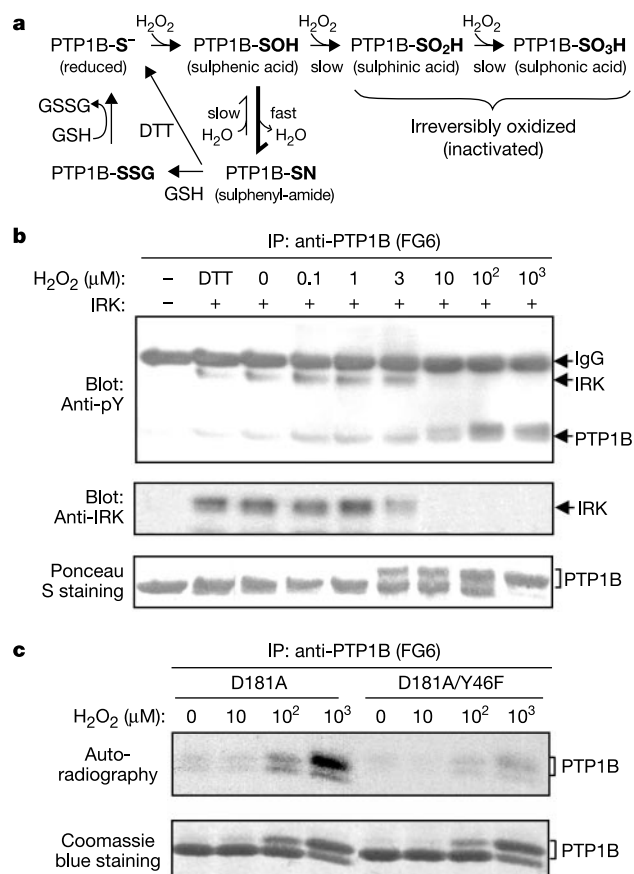


**Figure 3** Pulse-chase analysis of oxidation of PTP1B in solution. **a**, SDS-PAGE gels of wild-type PTP1B (left) and C215S PTP1B (right) incubated with increasing concentrations of  $H_2O_2$ , and the corresponding MALDI mass spectra for the Cys 215- and Ser 215-containing peptides after trypsinolysis. **b**, Left, activity of PTP1B after incubation with various concentrations of  $H_2^{18}O_2$  and associated recovery. Right, the equivalent  $M_r$  spectra of the Cys215-containing peptide obtained after chasing with a 100-fold excess of  $H_2^{18}O_2$  followed by trypsinolysis.

the insulin receptor (IRK), a physiological substrate of PTP1B<sup>17,18</sup>, and  $H_2O_2$ . We found that at concentrations of  $H_2O_2$  that resulted in PTP1B oxidation, binding of PTP1B to IRK was disrupted, consistent with inhibition of substrate recognition by oxidation (Fig. 4b). At these higher concentrations of  $H_2O_2$ , which did not impair IRK autophosphorylation activity, tyrosine phosphorylation of PTP1B by IRK was enhanced (Fig. 4b and Supplementary Information).

Consistent with the oxidation-induced structural changes in PTP1B, in which Tyr 46 adopts a solvent-exposed position, we noted that phosphorylation of a PTP1B mutant, in which Tyr 46 was changed to phenylalanine, was attenuated (Fig. 4c). Such redox-induced phosphorylation and changes in substrate recognition suggest that there are additional tiers of control over the regulation of PTP1B-mediated signalling pathways<sup>19</sup>.

It has been proposed that the catalytic cysteines of Cdc25, the low- $M_r$  PTPs, and the tumour suppressor phosphatase PTEN are protected from irreversible oxidation by disulphide bond formation with vicinal cysteines<sup>20,21</sup>; however, the tyrosine-specific PTPs do not contain equivalent vicinal thiols at their active sites. Here we have described an alternative mechanism to prevent irreversible oxidation, which also facilitates formation of glutathionylated derivatives of PTP1B<sup>22</sup>, thereby allowing redox-dependent regulation of these enzymes (Fig. 4a). Our findings of a previously unknown oxidation-dependent post-translational modification,



**Figure 4** Effects of oxidation on substrate binding and phosphorylation of PTP1B. **a**, Interrelationship of PTP1B redox species. **b**, Effects of oxidation on binding of the kinase domain of the insulin receptor (IRK) by D181A PTP1B. Immunoblot analysis of PTP1B immunoprecipitates with antibodies against pTyr and IRK shows that increasing concentrations of  $H_2O_2$  lead to disruption of a PTP1B-IRK complex and that tyrosine phosphorylation of PTP1B by IRK is enhanced. Ponceau S staining of the filter confirmed that an equal amount of PTP1B was immunoprecipitated from each sample (bottom). **c**, Mutation of Tyr 46 in PTP1B leads to attenuation of oxidation-induced phosphorylation of the PTP by IRK. Tyrosine phosphorylation of PTP1B D181A and D181A/Y46F in the IRK assay described above is compared in the autoradiograph. The Coomassie-stained gel is shown below.

coupled to an allosteric transition of PTP1B, will have implications for understanding mechanisms of redox signalling that are important in the regulation of gene expression and signal transduction<sup>1-4,23,24</sup>.

## Methods

### Oxidized crystals of PTP1B and crystallization

We incubated 32  $\mu M$  PTP1B in DTT-free buffer (10 mM Tris (pH 7.5), 25 mM NaCl and 0.2 mM EDTA) with 40  $\mu M$   $H_2O_2$  (ratio of 1:1.25) for 50 min at room temperature (20 °C) and crystallized<sup>13</sup> it without DTT. Data were collected on crystals at 3 weeks and 3 months after the protein was put into crystallization trials. The structures were solved as described below.

### Time course of oxidation

We washed PTP1B crystals for 1–2 min in 40  $\mu l$  of crystallization solution<sup>13</sup> to remove DTT, transferred them to 40  $\mu l$  of the same solution plus 50  $\mu M$   $H_2O_2$  and soaked them for durations of 20 min to 16 h (Supplementary Table 1) before freezing them at 100 K. Data were collected either in-house or at the Synchrotron Radiation Source (SRS, Daresbury, UK) or the European Synchrotron Radiation Facility (ESRF, Grenoble, France) and were processed with the HKL<sup>25</sup> package (Supplementary Information). The resolution ranged from 2.3 to 1.7 Å. We used apo-PTP1B as a starting model for refinement<sup>13</sup> by CNS<sup>26</sup>. Final structures were obtained by an additional 1–2 rounds of refinement in REFMAC<sup>29</sup> with the appropriate monomer libraries (Supplementary Table 1).

## Reduction of the sulphenyl-amide bond

Crystals prepared as above were transferred into 4  $\mu$ l of crystallization buffer containing 20  $\mu$ M H<sub>2</sub>O<sub>2</sub> and soaked for 12 h (before reduction with DTT) or 16 h (before reduction with GSH). After the H<sub>2</sub>O<sub>2</sub> soak, one crystal was frozen for data collection to verify the sulphenyl-amide state; the other three crystals were incubated in 40  $\mu$ l of crystallization buffer with either DTT (5 mM) or GSH (5 mM) for 72 h before data collection.

## Analysis of PTP1B in solution by mass spectrometry

PTP1B or the C215S mutant (56  $\mu$ M) was incubated with the indicated concentrations of H<sub>2</sub>O<sub>2</sub> for 10 min in 50 mM HEPES (pH 7.5), 1 mM EDTA and 200  $\mu$ M DTT. One microlitre of this solution was then digested with trypsin in 50 mM ammonium bicarbonate (pH 8.0) and analysed by MALDI-TOF mass spectrometry, and 10  $\mu$ l were boiled for 10 min in SDS loading buffer and analysed by SDS polyacrylamide gel electrophoresis (SDS-PAGE).

To reversibly oxidize PTP1B (pulse-chase experiment), we added 30  $\mu$ l of 'heavy' H<sub>2</sub><sup>18</sup>O<sub>2</sub> (2% stock solution, Icon 19) to an equal volume of 26  $\mu$ M PTP1B in degassed buffer (50 mM HEPES (pH 7.5), 1 mM EDTA and 100  $\mu$ M DTT) to give final concentrations of between 5 and 5  $\times$  10<sup>4</sup>  $\mu$ M. At 30 min, 10  $\mu$ l of PTP1B mixture was removed, chased with light H<sub>2</sub><sup>16</sup>O<sub>2</sub> (>100-fold excess) for 2 h and then trypsinized. We used 20  $\mu$ l of the solution to assay enzyme activity (% inhibition) with *para*-nitrophenol phosphate as substrate. The remaining 30  $\mu$ l of sample was incubated with DTT (10 mM) and catalase (10 U) for 2 h, and the recovered enzyme activity was measured (% recovery).

For trypsinolysis, PTP1B was diluted with 50 mM ammonium bicarbonate (pH 8.0) and 5 mM DTT, and incubated with 0.3  $\mu$ g of trypsin at 37 °C for 12–16 h. We mixed 1  $\mu$ l of this digest with saturated  $\alpha$ -cyano-4-hydroxycinnamic acid and analysed it with a Voyager DE-RP mass spectrometer (Applied Biosystems). Internal reference standards were used for calibration.

## Oxidation-induced disruption of PTP1B-IRK

We oxidized 12  $\mu$ g of the PTP1B substrate-trapping mutant D181A (ref. 16) or D181/Y46F in 50 mM Tris (pH 7.6) and 0.1 mM DTT with increasing concentrations of H<sub>2</sub>O<sub>2</sub> at room temperature (20 °C) in a total volume of 400  $\mu$ l. After 10 min, kinase buffer (50 mM Tris (pH 7.6), 4 mM MnCl<sub>2</sub>, 10 mM MgCl<sub>2</sub> and 100  $\mu$ M ATP), purified autophosphorylated IRK (2  $\mu$ g) and [ $\gamma$ -<sup>32</sup>P]ATP were added to each sample (1 ml final volume). After 20 min, we terminated the reaction by adding EDTA, immunoprecipitated PTP1B with FG6 antibody, resolved the proteins by SDS-PAGE and analysed them with antibodies against pTyr and against IRK.

Received 29 January; accepted 7 April 2003; doi:10.1038/nature01680.

- Finkel, T. Oxygen radicals and signalling. *Curr. Opin. Cell Biol.* **10**, 248–253 (1998).
- Finkel, T. Oxidants painting the cysteine chapel: redox regulation of PTPs. *Dev. Cell* **2**, 251–259 (2002).
- Rhee, S. G., Bae, Y. S., Lee, S. R. & Kwon, J. Hydrogen peroxide: a key messenger that modulates protein phosphorylation through cysteine oxidation. *Sci. STKE* **53** (2000); available at ([http://www.stke.org/cgi/content/full/OC\\_sigtrans;2000/53/pe1](http://www.stke.org/cgi/content/full/OC_sigtrans;2000/53/pe1)).
- Sundaresan, M., Yu, Z. X., Ferrans, V. J., Irani, K. & Finkel, T. Requirement for generation of H<sub>2</sub>O<sub>2</sub> for platelet-derived growth factor signal transduction. *Science* **270**, 296–299 (1995).
- Rao, G. N. Hydrogen peroxide induces complex formation of SHC–Grb2–SOS with receptor tyrosine kinase and activates Ras and extracellular signal-regulated protein kinases group of mitogen-activated protein kinases. *Oncogene* **13**, 713–719 (1996).
- Bae, Y. S. *et al.* Epidermal growth factor (EGF)-induced generation of hydrogen peroxide. Role in EGF receptor-mediated tyrosine phosphorylation. *J. Biol. Chem.* **272**, 217–221 (1997).
- Knebel, A., Rahmsdorf, H. J., Ullrich, A. & Herrlich, P. Dephosphorylation of receptor tyrosine kinases as target of regulation by radiation, oxidants or alkylating agents. *EMBO J.* **15**, 5314–5325 (1996).
- Mahadev, K., Zilbering, A., Zhu, L. & Goldstein, B. J. Insulin-stimulated hydrogen peroxide reversibly inhibits protein-tyrosine phosphatase 1B *in vivo* and enhances the early insulin action cascade. *J. Biol. Chem.* **276**, 21938–21942 (2001).
- Lee, S. R., Kwon, K. S., Kim, S. R. & Rhee, S. G. Reversible inactivation of protein-tyrosine phosphatase 1B in A431 cells stimulated with epidermal growth factor. *J. Biol. Chem.* **273**, 15366–15372 (1998).
- Barford, D., Das, A. K. & Egloff, M.-P. The structure and mechanism of protein phosphatases: insights into catalysis and regulation. *Annu. Rev. Biophys. Biomol. Struct.* **27**, 133–164 (1998).
- Denu, J. M. & Tanner, K. G. Specific and reversible inactivation of protein tyrosine phosphatases by hydrogen peroxide: evidence for a sulfenic acid intermediate and implications for redox regulation. *Biochemistry* **37**, 5633–5642 (1998).
- Collaborative Computational Project No. 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
- Barford, D., Flint, A. J. & Tonks, N. K. Crystal structure of human protein tyrosine phosphatase 1B. *Science* **263**, 1397–1404 (1994).
- Jia, Z., Barford, D., Flint, A. J. & Tonks, N. K. Structural basis for phosphotyrosine peptide recognition by protein tyrosine phosphatase 1B. *Science* **268**, 1754–1758 (1995).
- Pannifer, A. D., Flint, A. J., Tonks, N. K. & Barford, D. Visualization of the cysteinyl-phosphate intermediate of a protein-tyrosine phosphatase by X-ray crystallography. *J. Biol. Chem.* **273**, 10454–10462 (1998).
- Flint, A. J., Tiganis, T., Barford, D. & Tonks, N. K. Development of 'substrate-trapping' mutants to identify physiological substrates of protein tyrosine phosphatases. *Proc. Natl Acad. Sci. USA* **94**, 1680–1685 (1997).
- Elchebly, M. *et al.* Increased insulin sensitivity and obesity resistance in mice lacking the protein tyrosine phosphatase-1B gene. *Science* **283**, 1544–1548 (1999).
- Klaman, L. D. *et al.* Increased energy expenditure, decreased adiposity and tissue-specific insulin sensitivity in protein-tyrosine phosphatase 1B-deficient mice. *Mol. Cell. Biol.* **20**, 5479–5489 (2000).
- Haj, F. G., Markova, B., Klaman, L. D., Bohmer, F. D. & Neel, B. G. Regulation of receptor tyrosine kinase signaling by protein tyrosine phosphatase-1B. *J. Biol. Chem.* **278**, 739–744 (2003).

- Caselli, A. *et al.* The inactivation mechanism of low molecular weight phosphotyrosine-protein phosphatase by H<sub>2</sub>O<sub>2</sub>. *J. Biol. Chem.* **273**, 32554–32560 (1998).
- Lee, S. R. *et al.* Reversible inactivation of the tumour suppressor PTEN by H<sub>2</sub>O<sub>2</sub>. *J. Biol. Chem.* **277**, 20336–20342 (2002).
- Barrett, W. C. *et al.* Regulation of PTP1B via glutathionylation of the active site cysteine 215. *Biochemistry* **38**, 6699–6705 (1999).
- Kim, S. O. *et al.* OxyR: a molecular code for redox-related signaling. *Cell* **109**, 383–396 (2002).
- Blanchetot, C., Tertoolen, L. G. & den Hertog, J. Regulation of receptor protein-tyrosine phosphatase alpha by oxidative stress. *EMBO J.* **15**, 493–503 (2002).
- Otwiniowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
- Brunger, A. T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).

**Supplementary Information** accompanies the paper on [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank H. Jhoti for communicating data before submission. This work was funded by grants from Cancer Research UK (D.B.) and the NIH (N.K.T.).

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to D.B. ([david.barford@icr.ac.uk](mailto:david.barford@icr.ac.uk)). The coordinates of the sulphenyl-amide and sulphonic acid species of PTP1B have been deposited in the Protein Data Bank under accession codes 1OEM and 1OEO, respectively.

# Oxidation state of the active-site cysteine in protein tyrosine phosphatase 1B

Rob L. M. van Montfort, Miles Congreve, Dominic Tisi, Robin Carr & Harren Jhoti

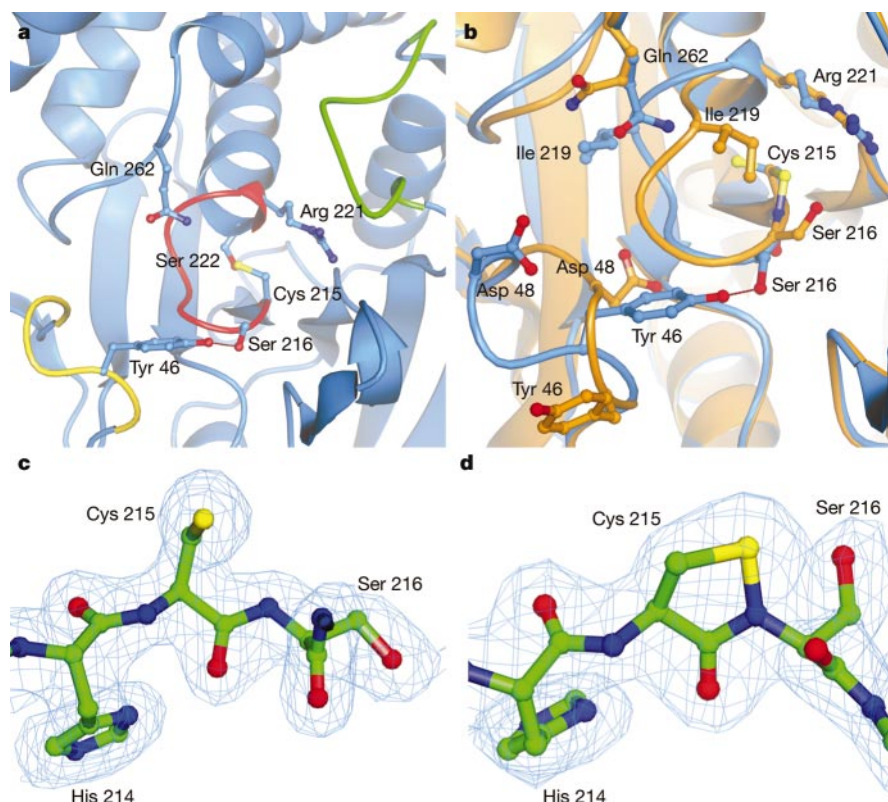
Astex Technology Ltd, 436 Cambridge Science Park, Milton Road, Cambridge CB4 0QA, UK

Protein tyrosine phosphatases regulate signal transduction pathways involving tyrosine phosphorylation<sup>1</sup> and have been implicated in the development of cancer, diabetes, rheumatoid arthritis and hypertension<sup>2</sup>. Increasing evidence suggests that the cellular redox state is involved in regulating tyrosine phosphatase activity through the reversible oxidation of the catalytic cysteine to sulphenic acid (Cys-SOH)<sup>3–6</sup>. But how further oxidation to the irreversible sulphinic (Cys-SO<sub>2</sub>H) and sulphonic (Cys-SO<sub>3</sub>H) forms is prevented remains unclear. Here we report the crystal structures of the regulatory sulphenic and irreversible sulphinic and sulphonic acids of protein tyrosine phosphatase 1B (PTP1B), an important enzyme in the negative regulation of the insulin receptor<sup>7,8</sup> and a therapeutic target in type II diabetes and obesity<sup>9</sup>. We also identify a sulphenyl-amide species that is formed through oxidation of its catalytic cysteine. Formation of the sulphenyl-amide causes large changes in the PTP1B active site, which are reversible by reduction with the cellular reducing agent glutathione. The sulphenyl-amide is a protective intermediate in the oxidative inhibition of PTP1B. In addition, it may facilitate reactivation of PTP1B by biological thiols and signal a unique state of the protein.

PTP1B belongs to a large family of protein tyrosine phosphatases (PTPs) characterized by an 11-residue signature sequence (I/V)HCXAGXXR(S/T/G), which includes the catalytic cysteine (Cys 215)<sup>1,10</sup>. Its catalytic mechanism involves a nucleophilic attack by Cys 215 on a phosphotyrosine (pTyr) substrate, resulting in a covalent phosphocysteine intermediate that is subsequently hydrolysed by an activated water molecule<sup>11</sup>.

The crystal structure of PTP1B shows that the PTP signature





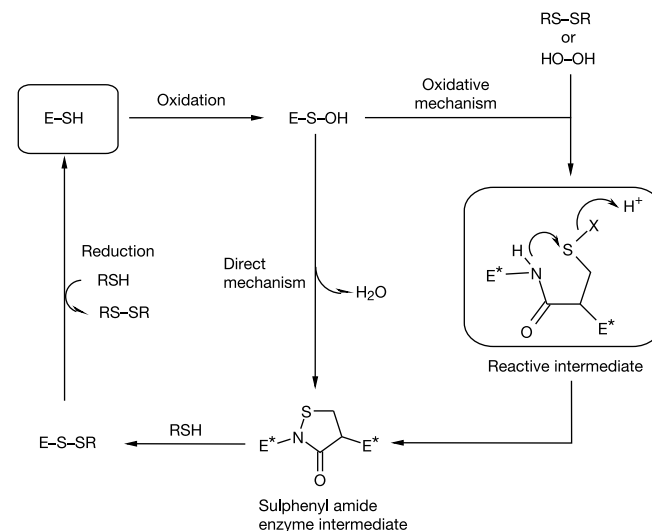
**Figure 1** Comparison of native and sulphenyl-amide PTP1B. **a**, Ribbon diagram of PTP1B. The phosphate-binding cradle is shown in red, the WPD loop in green and the pTyr recognition loop in gold. **b**, Superposition of native PTP1B (blue) and the sulphenyl-amide-containing structure (orange), showing different conformations of the pTyr recognition loop and the phosphate-binding cradle. **c**, Electron density of the catalytic cysteine and its

neighbouring residues in reduced PTP1B (see Supplementary Information). **d**, Electron density of the newly identified sulphenyl-amide derivative of Cys 215. The electron density maps in **c** and **d** are contoured at  $1\sigma$ . All figures are generated using Aesop (M. Noble, Laboratory of Molecular Biophysics, University of Oxford, unpublished).

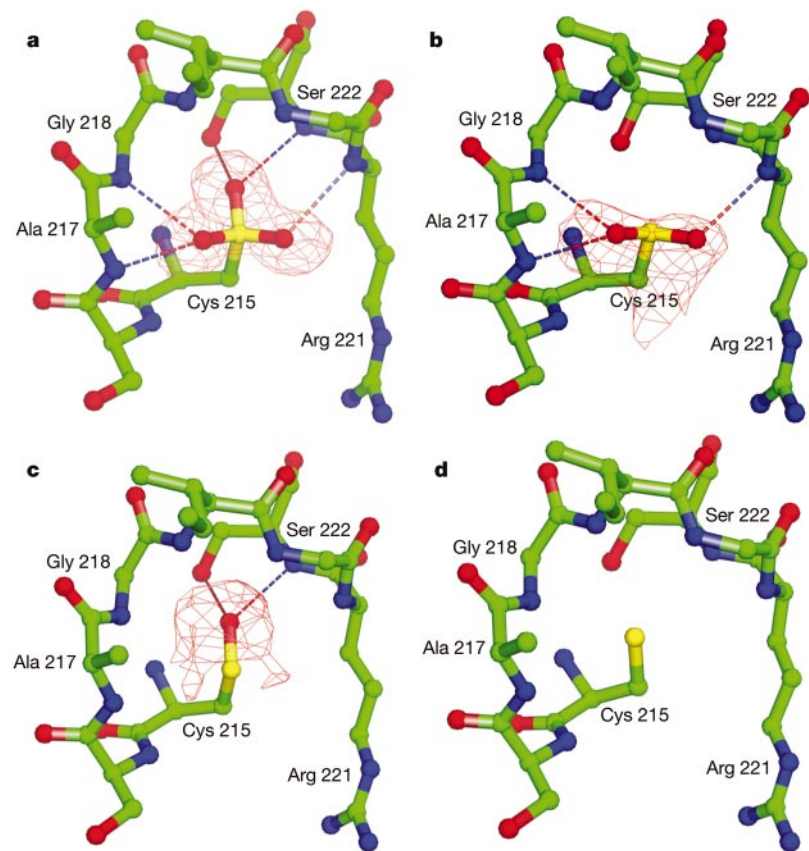
motif adopts a cradle-like conformation forming the base of the active site (ref. 12 and Fig. 1a). Its backbone amide atoms point to the centre of the cradle, which, together with the invariant Arg 221, provides an excellent environment to stabilize the negatively charged Cys 215 side chain and to bind the phosphate moiety of an incoming substrate (ref. 13 and Fig. 1a). A hydrogen bond between the hydroxyl group of Ser 222 and the sulphur atom of Cys 215 further stabilizes the cysteine conformation and helps to maintain its reduced  $pK_a$  ( $\sim 5.4$ )<sup>14</sup>.

One side of the active site is flanked by the so-called WPD loop, which adopts an open conformation in the unliganded enzyme and closes over a bound pTyr residue<sup>13</sup>. The opposite side is formed by the pTyr recognition loop containing Tyr 46, which mainly determines the specificity for pTyr substrates (ref. 15 and Fig. 1a). After carrying out an enzyme assay screen for potential inhibitors of PTP1B, we probed the oxidation state of the catalytic cysteine of PTP1B as we thought that it might be the cause of the high percentage of false-positive hits identified in the screen. Here we describe the crystal structures of a previously unknown sulphenyl-amide intermediate of PTP1B and sulphenic, sulphinic and sulphonic PTP1B derivatives, which we observed on exposing PTP1B crystals to potential inhibitors identified in the enzyme assay.

A crystallographic soaking experiment with 2-phenyl-isoxazolidine-3,5-dione, one of the identified PTP1B inhibitors, detected unexpected electron density close to the side chain of the catalytic cysteine that could be explained only by the presence of a covalent bond between the sulphur  $S_\gamma$  atom of Cys 215 and the backbone nitrogen atom of Ser 216 (Fig. 1c, d). This sulphenyl-amide bond has a bond length of 1.7 Å and results in a five-membered puckered ring that has not been previously observed in proteins.



**Figure 2** Putative mechanism of sulphenyl-amide formation and subsequent reactivation. The catalytic cysteine of PTP1B (E-SH) is oxidized to a sulphenic acid (E-S-OH). The sulphenyl-amide may be formed by a direct mechanism involving a nucleophilic attack of the backbone nitrogen of Ser 216 on the  $S_\gamma$  atom of Cys 215 and subsequent release of water. Alternatively, the sulphenic acid may be oxidized to a highly reactive intermediate by  $H_2O_2$  or an oxidized thiol, which then reacts to give the sulphenyl-amide. Reactivation of the enzyme occurs via mixed disulphide formation with a thiol. R, glutathione or DTT; X, leaving group OOH (sulphenoperoxy acid) or OS(O)R (sulphinothioic acid).



**Figure 3** Different oxidation states of the catalytic cysteine. Shown are structures of the sulphenic (a), sulphinic (b), sulphenic (c) acid derivatives and the reduced state (d) of Cys 215. The phosphate-binding cradle comprising residues 215–222 is shown in ball-and-stick representation using the same colour scheme as in Fig. 1c and d. Hydrogen bonds

are shown as broken lines. The  $\sigma_A$ -weighted  $F_o - F_c$  electron density map revealing the oxygen atoms is shown in red. The maps are contoured at  $3\sigma$  and in all maps the peaks are higher than  $5\sigma$ .

In conjunction with the formation of this sulphenyl-amide derivative, the phosphate-binding cradle adopts a distorted conformation, which is similar to, but not as pronounced as, that seen in the structure of the inactive C215S PTP1B mutant<sup>16</sup>. The cradle has shifted into the pTyr-binding site and stabilizes the sulphenyl-amide by a hydrophobic interaction with the side chain of Ile 219 (Fig. 1b). In addition, the side chain of Gln 262 moves out of the active site and also the pTyr loop adopts a unique conformation (Fig. 1b). The more exposed conformation of the pTyr loop results from the loss of the hydrogen bond between the hydroxyl groups of

Tyr 46 and Ser 216, which anchors the pTyr loop in native PTP1B and is stabilized by a network of water molecules mediating interactions between Asp 48 and the rest of the protein.

Formation of the sulphenyl-amide bond most probably occurs through oxidation of Cys 215 to sulphenic acid, followed by a nucleophilic attack of the backbone nitrogen atom of Ser 216 on the S $\gamma$  atom of Cys 215. Indeed, it has been postulated that the hydrogen bond interaction between the carbonyl oxygen atom of Cys 215 and the N $\delta$ 1 atom of the invariant His 214 side chain in native PTP1B increases the partial charge on the backbone nitrogen atom of

**Table 1** Crystallographic data collection and refinement statistics

|                                    | Sulphenic acid | Sulphinic acid | Sulphenic acid | Sulphenyl-amide | Back-soaking experiment |              |
|------------------------------------|----------------|----------------|----------------|-----------------|-------------------------|--------------|
|                                    |                |                |                |                 | Compound                | Glutathione  |
| Data collection                    |                |                |                |                 |                         |              |
| Beamline                           | In-house       | In-house       | In-house       | SRS 14.1        | In-house                | In-house     |
| $\lambda$ (Å)                      | 1.54           | 1.54           | 1.54           | 1.488           | 1.54                    | 1.54         |
| Resolution (Å)                     | 2.3            | 2.6            | 2.2            | 2.2             | 2.4                     | 2.2          |
| No. of observations                | 43,942         | 29,624         | 53,452         | 83,434          | 103,350                 | 108,793      |
| No. of unique reflections          | 20,755         | 16,237         | 23,367         | 23,591          | 18,514                  | 23,322       |
| Completeness (%)                   | 96.0 (94.9)    | 95.9(95.3)     | 96.0 (86.9)    | 98.9 (97.4)     | 99.8 (100)              | 100 (100)    |
| $R_{\text{merge}}$                 | 0.092 (0.25)   | 0.172 (0.37)   | 0.060 (0.252)  | 0.050 (0.26)    | 0.078 (0.273)           | 0.054 (0.32) |
| $I/\sigma(I)$                      | 7.3 (2.8)      | 4.6 (1.9)      | 11.2 (4.3)     | 7.8 (2.8)       | 8.4 (2.8)               | 12.9 (2.2)   |
| Refinement                         |                |                |                |                 |                         |              |
| $R_{\text{cryst}}/R_{\text{free}}$ | 0.168/0.226    | 0.196/0.270    | 0.172/0.217    | 0.182/0.233     | 0.215/0.275             | 0.228/0.277  |
| r.m.s.d. bond lengths (Å)          | 0.013          | 0.018          | 0.012          | 0.009           | 0.006                   | 0.006        |
| r.m.s.d. bond angles (°)           | 1.4            | 1.7            | 1.3            | 1.4             | 1.3                     | 1.3          |

Numbers in parentheses indicate the highest shell values.  $R_{\text{merge}} = \sum_h \sum_i |I(h,i) - \langle I(h) \rangle| / \sum_h \sum_i I(h,i)$ , where  $I(h,i)$  is the scaled intensity of the  $i$ th observation of reflection  $h$  and  $\langle I(h) \rangle$  is the mean value. Summation is over all measurements.  $R_{\text{cryst}} = \sum_{hkl} |F_o| - k|F_c| / \sum_{hkl} |F_o|$ , where  $F_o$  and  $F_c$  are the observed and calculated structure factors,  $k$  is a weighting factor and work denotes the working set of 95% of the reflections used in the refinement.  $R_{\text{free}} = \sum_{hkl} |F_o| - k|F_c| / \sum_{hkl} |F_o|$ , where test denotes the test set of 5% of the reflections used in cross-validation of the refinement.  $\lambda$  refers to wavelength, r.m.s.d. to root mean square deviations.

Ser 216 (ref. 12), enhancing its reactivity and supporting a nucleophilic substitution mechanism.

*In vivo* the sulphenic acid can be formed by oxidation with hydrogen peroxide ( $\text{H}_2\text{O}_2$ )<sup>17</sup>, but under our experimental conditions it is most probably triggered by redox cycling of 2-phenylisoxazolidine-3,5-dione (Fig. 2). The leaving group in the cyclization reaction could be a water molecule in an  $\text{S}_\text{N}2$  substitution reaction (Fig. 2, direct mechanism). Alternatively, the sulphenic acid derivative might be oxidized under the experimental soaking conditions, or by physiological oxidants such as  $\text{H}_2\text{O}_2$  or oxidized glutathione *in vivo*, to form a highly reactive intermediate. This would result in faster formation of the sulphenyl-amide species (Fig. 2, oxidative mechanism).

Sulphenic acids are generally labile and can be easily oxidized further to their sulphinic or sulphonic forms<sup>17–19</sup>. Because of the high reactivity of Cys 215, we were not surprised to find it oxidized to either sulphonic (Fig. 3a) or sulphinic acid (Fig. 3b) in some crystallographic soaking experiments with putative inhibitors. Under two different sets of soaking conditions, however, we were able to visualize the sulphenic acid derivative by crystallography (Fig. 3c). We think that the lifetime of this intermediate species may be extended by the mild oxidizing agents used and could be influenced by the pH of the experimental conditions. In contrast to the sulphenyl-amide derivative, the structures of these oxidation states are very similar to the structure of reduced PTP1B.

In the sulphonic acid PTP1B derivative, one of the oxygen atoms in the modified Cys 215 forms a hydrogen bond with the backbone amide of Arg 221; the second oxygen atom is stabilized by the backbone amide groups of Ala 217 and Gly 218 and a water molecule, which also interacts with the side chain of Gln 262 and the amide group of Gly 220; and the third oxygen atom forms hydrogen bonds with both the amide and the hydroxyl groups of Ser 222.

The sulphinic acid intermediate is stabilized by the same interactions observed for the first two oxygen atoms in the sulphonic acid derivative. Intriguingly, the interactions of the sulphenic acid intermediate are identical to the ones made by the third sulphonic oxygen atom, which forms hydrogen bonds with both the main chain and side chain of Ser 222. Consequently, the hydrogen bond interaction between the Cys 215 side chain and the hydroxyl group of Ser 222, which is present in native PTP1B (Fig. 3d) and thought to be important in hydrolysis of the phosphocysteine intermediate<sup>14</sup>, has been disrupted. Thus, the sulphenic acid derivative not only blocks the formation of the phosphocysteine intermediate, but also inhibits the second step of the reaction, indicating its effectiveness in inactivating the enzyme.

PTP1B can be reactivated by reduction with dithiothreitol (DTT) or glutathione, which involves the formation of mixed disulphides with Cys 215 (ref. 3). But the Cys-SOH structure shows that although the sulphenic  $\text{S}_\gamma$  atom is accessible, its hydroxyl group is almost completely buried at the bottom of the active site. Its exit as a water molecule will be severely impeded on disulphide formation with a thiol, and therefore reactivation of the sulphenic acid intermediate in this conformation seems unlikely. However, a nucleophilic attack of the nearby backbone nitrogen of Ser 216, coupled to the conformational changes observed in the sulphenyl-amide derivative, would generate sufficient space for a water molecule to leave the active site.

In addition, the  $\text{S}_\gamma$  atom in the sulphenyl-amide intermediate occupies a position that is more accessible for reduction by biological thiols. Alternatively, to account for the fact that the sulphenic acid derivative can be trapped, it is possible that further oxidation of the cysteine might be required to catalyse the formation of the sulphenyl-amide species (Fig. 2, oxidative mechanism). Such a further oxidative step might allow an extra level of regulatory control by cellular oxidants under physiological conditions.

Although we can only speculate about an *in vivo* role for the sulphenyl-amide intermediate, our results suggest that it functions in the oxidative regulation of PTP1B by preventing the irreversible oxidation of Cys 215 and by facilitating its thiol-mediated reactivation (Fig. 2). To verify the reversibility of the S–N bond, we attempted to reduce the sulphenyl-amide derivative in the crystals. Two PTP1B crystals were soaked in a solution containing 2-phenylisoxazolidine-3,5-dione. X-ray data collected from one of the crystals confirmed the formation of the sulphenyl-amide bond and the concomitant conformational changes in the active site. The second crystal was back-soaked in 20 mM reduced glutathione in an attempt to reduce the Cys 215 sulphenyl-amide derivative back to its native form. X-ray data collected from the back-soaked crystal showed that the active site had returned to its native conformation, thereby confirming structurally the reactivation of the sulphenyl-amide PTP1B derivative by a biologically relevant reducing agent and strengthening the hypothesis that the sulphenyl-amide has a protective role in PTP1B redox regulation.

In conclusion, our studies provide a detailed structural understanding of the intermediates involved in redox regulation of PTP1B and identify a previously unknown oxidation state of its catalytic cysteine. The formation of the sulphenyl-amide intermediate is an elegant mechanism to protect Cys 215 from further oxidation, and the concomitant conformational changes of the phosphate-binding cradle and pTyr loop may signal the inactive state of the enzyme. The structures of the sulphenic acid and sulphenyl-amide derivative indicate that reactivation of PTP1B seems to be facilitated by the sulphenyl-amide form. A better understanding of the mechanism of inhibition of PTP1B should aid the design of potential inhibitors against this key therapeutic target for treating diabetes. □

## Methods

### PTP1B purification and crystallization

Expression, purification and crystallization of the catalytic domain of PTP1B (residues 1–321) were based on described conditions<sup>20</sup>. For overnight soaking experiments, we transferred crystals into a standard mother liquor containing 16% PEG 4000, 0.1 M HEPES (pH 7.5), 0.2 M magnesium acetate, 10 mM DTT and the compound of interest (see Supplementary Information for details).

### Structure determination

All data were collected at 100 K. Data sets of sulphenic, sulphinic and sulphonic acid PTP1B derivatives were automatically collected on a Jupiter140 CCD (charge-coupled device) detector mounted on a RU3HR rotating anode generator equipped with an ACTOR sample-changing robot (all RigakuMSC). We processed and scaled them with D\*TREK<sup>21</sup> and then converted them to structure factors by using programs from the CCP4 suite<sup>22</sup>. Oxidation of Cys 215 to the sulphenyl-amide derivative was done by soaking crystals for about 24 h in mother liquor without DTT but containing 100 mM 2-phenylisoxazolidine-3,5-dione. Data were collected at station 14.1 (Synchrotron Radiation Source, Daresbury, UK) and processed and scaled using MOSFLM/SCALA<sup>22</sup>.

We checked the reversibility of the sulphenyl-amide derivative by first soaking two crystals in 100 mM 2-phenylisoxazolidine-3,5-dione for 24 h. Sulphenyl-amide formation was confirmed by an in-house data set collected on one of the crystals. After back-soaking for 24 h in mother liquor containing 20 mM reduced glutathione, data collected from the other crystal confirmed the active site in its native conformation. All results were evaluated using the graphics programs QUANTA (Accelrys) and Astexviewer<sup>23</sup>. Data collection statistics are summarized in Table 1. Initial refinement was always carried out using the CCP4-based Astex automatic refinement scripts, followed by rounds of positional and B-factor refinement with REFMAC5 (ref. 22), alternated with manual rebuilding steps using QUANTA. We generated monomer libraries for the different oxidation states by using REFMAC5 in combination with the monomer sketcher of the CCP4GUI (ref. 22).

Received 29 January; accepted 7 April 2003; doi:10.1038/nature01681.

1. Neel, B. G. & Tonks, N. K. Protein tyrosine phosphatases in signal transduction. *Curr. Opin. Cell Biol.* **9**, 193–204 (1997).
2. Zhang, Z. Y. Protein tyrosine phosphatases: structure and function, substrate specificity, and inhibitor development. *Annu. Rev. Pharmacol. Toxicol.* **42**, 209–234 (2002).
3. Denu, J. M. & Tanner, K. G. Specific and reversible inactivation of protein tyrosine phosphatases by hydrogen peroxide: evidence for a sulphenic acid intermediate and implications for redox regulation. *Biochemistry* **37**, 5633–5642 (1998).
4. Denu, J. M. & Dixon, J. E. Protein tyrosine phosphatases: mechanisms of catalysis and regulation. *Curr. Opin. Chem. Biol.* **2**, 633–641 (1998).
5. Meng, T. C., Fukada, T. & Tonks, N. K. Reversible oxidation and inactivation of protein tyrosine phosphatases *in vivo*. *Mol. Cell* **9**, 387–399 (2002).



6. Lee, S. R., Kwon, K. S., Kim, S. R. & Rhee, S. G. Reversible inactivation of protein-tyrosine phosphatase 1B in A431 cells stimulated with epidermal growth factor. *J. Biol. Chem.* **273**, 15366–15372 (1998).
7. Elchebly, M. *et al.* Increased insulin sensitivity and obesity resistance in mice lacking the protein tyrosine phosphatase-1B gene. *Science* **283**, 1544–1548 (1999).
8. Klamann, L. D. *et al.* Increased energy expenditure, decreased adiposity, and tissue-specific insulin sensitivity in protein-tyrosine phosphatase 1B-deficient mice. *Mol. Cell. Biol.* **20**, 5479–5489 (2000).
9. Goldstein, B. J. Protein-tyrosine phosphatase 1B (PTP1B): a novel therapeutic target for type 2 diabetes mellitus, obesity and related states of insulin resistance. *Curr. Drug. Targets Immune Endocr. Metabol. Disord.* **1**, 265–275 (2001).
10. Barford, D. Protein phosphatases. *Curr. Opin. Struct. Biol.* **5**, 728–734 (1995).
11. Pannifer, A. D., Flint, A. J., Tonks, N. K. & Barford, D. Visualization of the cysteinyl-phosphate intermediate of a protein-tyrosine phosphatase by X-ray crystallography. *J. Biol. Chem.* **273**, 10454–10462 (1998).
12. Barford, D., Flint, A. J. & Tonks, N. K. Crystal structure of human protein tyrosine phosphatase 1B. *Science* **263**, 1397–1404 (1994).
13. Jia, Z., Barford, D., Flint, A. J. & Tonks, N. K. Structural basis for phosphotyrosine peptide recognition by protein tyrosine phosphatase 1B. *Science* **268**, 1754–1758 (1995).
14. Lohse, D. L., Denu, J. M., Santoro, N. & Dixon, J. E. Roles of aspartic acid-181 and serine-222 in intermediate formation and hydrolysis of the mammalian protein-tyrosine-phosphatase PTP1. *Biochemistry* **36**, 4568–4575 (1997).
15. Sarmiento, M., Zhao, Y., Gordon, S. J. & Zhang, Z. Y. Molecular basis for substrate specificity of protein-tyrosine phosphatase 1B. *J. Biol. Chem.* **273**, 26368–26374 (1998).
16. Scapin, G., Patel, S., Patel, V., Kennedy, B. & Asante-Appiah, E. The structure of apo protein-tyrosine phosphatase 1B C215S mutant: more than just an S → O change. *Protein Sci.* **10**, 1596–1605 (2001).
17. Mahadev, K., Zilbering, A., Zhu, L. & Goldstein, B. J. Insulin-stimulated hydrogen peroxide reversibly inhibits protein-tyrosine phosphatase 1b *in vivo* and enhances the early insulin action cascade. *J. Biol. Chem.* **276**, 21938–21942 (2001).
18. Claiborne, A. *et al.* Protein-sulfenic acids: diverse roles for an unlikely player in enzyme catalysis and redox regulation. *Biochemistry* **38**, 15407–15416 (1999).
19. Claiborne, A., Miller, H., Parsonage, D. & Ross, R. P. Protein-sulfenic acid stabilization and function in enzyme catalysis and gene regulation. *FASEB J.* **7**, 1483–1490 (1993).
20. Barford, D., Kelly, J., Flint, A. J. & Tonks, N. K. Purification and crystallization of the catalytic domain of human protein tyrosine phosphatase 1B expressed in *Escherichia coli*. *J. Mol. Biol.* **239**, 726–730 (1994).
21. Pflugrath, J. W. The finer things in X-ray diffraction data collection. *Acta Crystallogr. D* **55**, 1718–1725 (1999).
22. Collaborative Computational Project No. 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
23. Hartshorn, M. J. A visualisation aid for structure-based drug design. *J. Comput. Aid. Mol. Des.* (in the press).

**Supplementary Information** accompanies the paper on [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank N. Wallis and G. Williams for discussions; O. Callaghan for compounds; I. Tickle for assistance with crystallographic software; and D. Barford for access to his data before submission and for discussions.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to H.J. ([h.jhoti@astex-technology.com](mailto:h.jhoti@astex-technology.com)). Coordinates of the sulphenyl-amide, sulphenic, sulphinic and sulphonic derivatives of PTP1B have been deposited in the Protein Data Bank under accession codes 1oes, 1oet, 1oeu and 1oev, respectively.

# Accept no substitute

New ideas, including an anti-counterfeiting kit.



Buffer state: 100 and 500 ml sizes from Amresco.

## A.C.E.

**Amresco** [www.amresco-inc.com](http://www.amresco-inc.com)

### Sequencing buffer comes up trumps

A.C.E. biotechnology grade sequencing buffer 1 × is an addition to Amresco's advanced capillary electrophoresis line. It is formulated for DNA sequencing via capillary electrophoresis and comes ready to use, with no dilution required. The buffer is suitable for use with the ABU 3100, 3700 and 310 automated CE systems as well as with the Beckman CEQ 2000XL sequencer.

## TurboCells *E. coli*

### Gene Therapy Systems

[www.genetherapysystems.com](http://www.genetherapysystems.com)

### Make a rapid recovery

TurboCells' chemically competent *E. coli* are optimized to yield transformation in just three minutes. Used together with the provided buffer, they eliminate the need for hour-long recovery following transformation. In addition, heat shock takes place at 37 °C, removing the inconvenience of a 42 °C water bath. TurboCells are available for both routine cloning and T7-promoter-based *E. coli* expression and are supplied in single-use 50 µl aliquots.

## Polystyrene containers

**Bibby Sterilin** [www.bibby-sterilin.com](http://www.bibby-sterilin.com)

### Have your fill

Bibby Sterilin's range of polystyrene containers have full compliance with IVD Directive 98/79/CE and bear the requisite CE mark. Sizes range from a 7 ml bijou through to a large 250 ml container. They may be bought unlabelled or with either plain or clinical labels, and come with a choice of plastic or metal flow seal caps. The range also includes boric acid versions of both the 30 ml universal

and the 7 ml bijou for urine collection and preservation, and a 30 ml option with an integral spoon for faecal sampling. For chlamydia culture, a 7 ml bijou 'trac' bottle containing a glass coverslip is available.

## MagPrep

### Tecan

[www.tecan.com](http://www.tecan.com)

### A winning combination for DNA extraction

MagPrep kits can be combined with Tecan's entry-level MiniPrep automated pipetting platform equipped with TeMagS robotic magnetic bead separation accessory to produce a cost-effective way of isolating and purifying DNA. The MagPrep genomic DNA kit purifies genomic DNA from fresh blood by capturing leukocytes on the surface of silica-coated magnetic beads. After removal of other cell types, the leukocytes are lysed, releasing the genomic DNA that adheres to the beads. Wash steps further separate the DNA from other cellular components. The released, soluble DNA is sheared to lengths of approximately 10,000 bp. Fragments from 100 bp to 10 KB can be purified using the MagPrep PCR Cleanup kit. Reagent use is scalable to 96 samples per run, optimizing throughput and minimizing cost.

## Orion II microplate luminometer

### Berthold Detection Systems

[www.berthold-ds.com](http://www.berthold-ds.com)

### Sophisticated features, simple to use

The Orion II microplate luminometer combines high-precision reagent injectors, optional temperature control and Windows 'Simplicity' software to provide advanced detection of luminescence from microplates. It can be supplied with up to four reagent injectors. Three modes are available for shaking of microplates. The unit has both RS-232 and USB serial ports, and supports robotic loading of microplates. An optional TestPlate is available for regular automatic validation of the



Orion II: detection on a microplate.

detector function. The TestPlate allows verification of the entire measuring range, covering up to six orders of magnitude, and produces a report according to GLP guidelines.

### DNA-tagged anti-counterfeit labels

**Applied DNA Sciences** [www.adnas.com](http://www.adnas.com)

#### *Protection for designer labels*

This new DNA-tagged label technology is a deterrent to the increasing number of sophisticated counterfeiters. It combines specially processed DNA with anti-counterfeit ink, which can be applied to paper, cloth or plastic labels or garment labels, and printed onto logos or any other surface. Labels are printed with proprietary ink containing a unique authentication DNA code for each manufacturer. The labels can be easily tested for manufacturers' authenticity and traced to the point of origin. In addition, labels can be instantly identified as containing DNA-tagged label technology imprinting with a second level of a proprietary embedded DNA. Virtually any item that can be duplicated can be protected. A simple real-time field test consists of a coded activation solution stored in a one-use swab. A simple wipe of the solution induces a predetermined colour change in the DNA-tagged ink used to print the label. This reaction only occurs in the presence of the DNA tag and cannot be replicated in a fake label product.

### Vulcan

**Advanced ChemTech**

[www.advancedchemtech.com](http://www.advancedchemtech.com)

#### *High-pressure speeds parallel organic synthesis*

The Vulcan high-pressure reactor provides a flexible tool for enhancing the productivity of automated organic syntheses. In conjunction with organic synthesizers such as the compa-

ny's Vantage system, it enables to 96 different and isolated reactions with volumes between 3.5 ml and 100 ml to be performed simultaneously. Reactions can be carried out with up to six different gases applied under high pressure (up to 105 bar) and across the temperature range  $-80^{\circ}\text{C}$  to  $+150^{\circ}\text{C}$ . The Vulcan accommodates reactions in solution or suspension where integral variable speed vortex mixing ensures consistent and sustainable reactions. Applications include catalyst development, catalyst screening and synthesis of new compounds at elevated pressure.

### GenPlate

**RNAture**

[www.rnature.com](http://www.rnature.com)

#### *Lyse, hybridize and go*

GenePlate's oligonucleotide immobilization technology allows mRNA isolation, cDNA synthesis and PCR to be performed in the same well. No columns, filters, beads, organic extractions or centrifugation are necessary. These versatile kits allow recovery of suspension cell mRNA, adherent cell mRNA and tissue mRNA. They are available in 96-well, 384-well and 8-well strip formats.

### Prelude Karl Fischer workstation

**Zymark**

[www.zymark.com](http://www.zymark.com)

#### *A solid investment*

This workstation automates sample handling and moisture determination of solid dosage forms, feeds and capsules for applications in the pharmaceutical, food, chemical, tobacco and paper industries. It uses the ThermoOrion Turbo2 titrator for complete moisture extraction from a sample into solution with little or no sample preparation. The system can handle a wide range of sample moisture levels with multiple sample analyses using the same sol-



vents, with a high-precision titrant delivery system to maximize accuracy and repeatability. Running 96 samples unattended, the Prelude Karl Fischer workstation employs easy-to-use software that is 21 CFR Part 11 compliant. The system also includes a barcode reader, a four-place Mettler SAG balance and a sample holder standard rack configuration.

### HIV 1/2 serum test

**Bionor** [www.bionor.no](http://www.bionor.no)

#### Speedy serology

Bionor has developed a simple and semi-rapid test for confirming HIV 1 and 2 serological status. The test is based on an ELISA principle

using magnetic beads as a solid support where antibody responses to individual viral proteins are visualized separately. Purified peptides to HIV 1 core p24<sup>CA</sup>, HIV-1 transmembrane glycoproteins gp41<sup>TM</sup>, HIV-1 envelope glycoproteins gp160<sup>ENV</sup> and HIV-2 transmembrane glycoproteins gp36<sup>TM</sup> are used as a source of antigen. The results are interpreted easily and serological status is determined according to the CDC/ASTPHLD guidelines. The use of peptides as a source of antigen minimizes non-specific responses that can often complicate the interpretation of western blot assays normally used to confirm HIV serological status. The Bionor test demonstrates equal sensitivity to current western blot assays and provides greater specificity due to fewer indeterminate results. By combining the Bionor rapid HIV 1 and 2 screening and confirmatory tests, HIV serological status can be detected and confirmed within hours.

### PI 3-kinase assay kit

**Echelon Biosciences** [www.echelon-inc.com](http://www.echelon-inc.com)

#### Non-radioactive, quick and sensitive

This assay kit provides rapid and sensitive measurement for *in vitro* PI 3-kinase activity. The kit is available in three formats: 96-well ELISA, 384-well fluorescence polarization, and 384-well AlphaScreen.

### OK Kit

**Microzone** [www.microzone.co.uk](http://www.microzone.co.uk)

#### Fitness check for thermal cyclers

Microzone's OK Kit is designed to test the performance of thermal cyclers. Inconsistent or unrepeatable PCR results may be a consequence of variations in heating temperatures, ramp rates, block performance and lid performance within and between thermal cyclers. The OK Kit is simple to use. Two reagents are mixed together, aliquoted into tubes or plates and placed in the thermal cycler. A simple cycling regime is undertaken and the PCR samples are electrophoresed on a standard agarose gel containing ethidium bromide. If the thermal cycler is performing correctly, three bands of different intensities will be clearly visible.

### ProtoCOL

**Synbiosis** [www.synbiosis.com](http://www.synbiosis.com)

#### Rapid quality assurance of vital vaccines

ProtoCOL is a reaction zone reader and colony counter that provides an indication of vaccine potency by automatically measuring reaction zones in single radial immunodiffusion (SRD) assays. The system consists of a CCD camera integrated with a PC and software. It can automatically read an entire SRD plate, measure reaction zones and transcribe the results in Excel. The software also flags areas of dispute such as fuzzy edges, so that users can measure them manually with the on-screen ruler. The



ProtoCOL: shedding light on vaccines.

results are produced in a flexible Excel format and can be exported directly into most other statistical programs.

### Custom-designed rotary valves

**Omnifit** [www.omnifit.com](http://www.omnifit.com)

#### They're bored rigid

Omnifit can design, prototype and manufacture manual and electric valves, bottle caps, tubing sets and chromatography columns for many different analytical and biotechnology instrumentation applications. The company's custom-designed valves are tailored to the user's specific requirements. One manufacturer of a DNA analysis instrument using high-viscosity fluids found that solenoid valves were ineffective, and needed an inert valve with sufficient force to cut through the fluids and act as a shut-off valve. Omnifit supplied a 1mm-bore custom-designed rotary valve with a stainless-steel rotor for torsional strength, a Tefzel liner for chemical inertness, and a PEEK body for rigidity.

### RNeasy MinElute

**Qiagen** [www.qiagen.com](http://www.qiagen.com)

#### Less is more

The RNeasy MinElute cleanup kit provides efficient cleanup of enzymatic reactions, such as labelling, and crude RNA preps in less than 15 minutes. The microspin assembly enables concentration of small amounts of RNA to only 10 µl. RNA amounts corresponding to less than one cell (as little as 1 pg) can be concentrated for use in sensitive downstream applications such as microarray analysis and real-time RT-PCR. Larger amounts (up to 45 mg) can be cleaned up and concentrated so that reaction volumes can be kept small, increasing reaction efficiency. The high purity allows the complete sample to be used in reactions, without inhibition.

*These notes are compiled in the Nature office from information provided by the manufacturers.*

## Contacts

**Publisher:** Ben Crowe

**Editor:** Paul Smaglik

**Marketing Manager:** David Bowen

### European Head Office, London

The Macmillan Building  
4 Crinan Street  
London N1 9XW, UK  
Tel +44 (0) 20 7843 4961  
Fax +44 (0) 20 7843 4996  
e-mail: [naturejobs@nature.com](mailto:naturejobs@nature.com)

**Senior European Sales Manager:**  
Nevin Bayoumi (4978)

### UK/ RoW/ Ireland:

Matt Powell (4953)

Andy Douglas (4975)

Frank Phelan (4944)

### Netherlands/ Italy/

Iberia/ Belgium:

Evelina Rubio Hakansson (4973)

**Scandinavia:** Silje Opstrup (4994)

**France/ Switzerland:**

Amelie Pequignot (4974)

**Production Manager:** Billie Franklin

To send materials use London  
address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: [naturejobs@nature.com](mailto:naturejobs@nature.com)

### International

**Advertising Coordinator:**

Hind Berrada (4935)

**Naturejobs web development:**

Tom Hancock

**Naturejobs online production:**

Ben Lund

### European Satellite Office

**Germany/ Austria:**

Patrick Phelan, Odo Wulffen

Tel + 49 89 54 90 57 11/-2

Fax + 49 89 54 90 57 20

e-mail: [p.phelan@nature.com](mailto:p.phelan@nature.com)

[o.wulffen@nature.com](mailto:o.wulffen@nature.com)

### US Head Office, New York

345 Park Avenue South,  
10th Floor, New York, NY 10010-1707  
Tel +1 800 989 7718

Fax +1 800 989 7103

e-mail: [naturejobs@natureny.com](mailto:naturejobs@natureny.com)

**US Sales Manager:** Peter Bless

**US Advertising Coordinator:**

Linda Adam

### Japan Head Office, Tokyo

MG Ichigaya Building (5F),  
19-1 Haraikatamachi,  
Shinjuku-ku,  
Tokyo 162-0841

Tel +81 3 3267 8751

Fax +81 3 3267 8746

**Asia-Pacific Sales Director:**

Hideki Watanabe

e-mail: [h.watanabe@naturejp.com](mailto:h.watanabe@naturejp.com)

# naturejobs

## Wrapped in red tape

In 1990, Rob Chess was an adviser to a Bush administration that expanded scientific immigration. Now, as chairman of biotech company Nektar Therapeutics in San Carlos, California, he is watching a different Bush administration contract it. Chess, like most scientists, understands the need for security in a changed world. But he bemoans its unintended consequences in the scientific community. The first ripples hit many foreign graduate students and postdocs when they tried to return to the United States after the 2002 Christmas holiday, delaying them and in some cases leaving them in limbo (see *Nature* **422**, 96–97; 2003). Then there were repercussions on broader academic hiring (see *Nature* **422**, 457–458; 2003).

Now, Chess sees signs that strangled immigration is constricting recruitment into industry. Rather than taking weeks to process immigration requests, it now takes 3–6 months, he says. This is problematic — with 60–65% of science and engineering PhDs in the United States granted to foreign scientists, firms have little choice but to seek skills abroad. “We’ve always recruited, even in tough times, the absolute best scientists, regardless of where they come from,” Chess says. “The best person in the world sometimes is not a US citizen.”

And the process seems to be getting more cumbersome. In May, the US Department of State instituted a policy to interview everyone applying for a visa. And last October, the number of H1B visas, which are often issued for faculty members, will be cut to 65,000 a year, from a high of 195,000. Even if those decreases hit academia hardest, they will have a knock-on effect to industry.

Unless the visa review process becomes more efficient, recruitment options for US companies wanting to be competitive with their global counterparts will shrink, along with the number of visas issued each year.

**Paul Smaglik**

*Naturejobs* editor



## Contents

### SPECIAL REPORT

The spread of  
epidemiology

p784

### WWW.NATUREJOBS.COM

Career centre  
Information on the  
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

ANNOUNCEMENTS

EVENTS



# With the United States on high alert over the possibility of bioterror attacks, epidemiologists are in huge demand, says Virginia Gewin.

**A**nthrax, West Nile virus, severe acute respiratory syndrome (SARS) and the increased threat of bioterrorism are just some of the factors that have increased the demand for epidemiologists recently — primarily in the United States. And the US government has not been slow to plough additional funds into the fight against bioterror, which in turn has created a number of epidemiology-related positions around the country.

But although the finances are in order, there is currently a significant deficit in skilled personnel to fill the new posts. “Now that states and localities have got bioterrorism funding, they are having a hard time hiring,” says Stephen Thacker, chief epidemiologist at the US Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia. In fact, the push to fill bioterror-related positions is drawing talent away from the public-health sector, aggravating an already strained healthcare system.

The Council of State and Territorial Epidemiologists (CSTE), a collective of US public-health professionals, estimates that at least 1,600 formally trained epidemiologists — a doubling of the current level — will be needed in the near future in response to growing public-health programmes and the increased focus on bioterrorism. To meet the demand, both federal and state funds are being applied to help generate fresh training opportunities.

Mike Osterholm would like to see more planning for the future.

## UNDER SURVEILLANCE

Leading the way is the CDC, which has several programmes designed to help those with medical or epidemiological training to become leaders in public health. Established in 1951, the CDC's Epidemic Intelligence Service programme is one of the most notable schemes. Lasting for two years, it gives fellows on-the-job training in surveillance and response units that deal with all sorts of epidemics including chronic disease, injuries and, now, bioterrorism.

In previous years, the programme took 60–70 out of 300 or so applicants, but this year's class of 80 was the largest ever, and was boosted by concerns about bioterrorism. In recent years, 15–20% of the class has been filled by fellows from other nations in the hope that these trainees would build up surveillance programmes in their home countries (see ‘European

approach’, below). At least 90% of the trainees end up in public-health jobs with either the CDC, the World Health Organization or state and local health departments. Increasingly, individual states are setting up similar training programmes to cultivate specialized epidemiological staff to meet state and local needs (see ‘Local knowledge’, opposite).

Indeed, now that more funds are available, there is a nationwide push to employ more support staff to allow health departments to deal with the increased volume of work and to improve responsiveness. “We want to develop the capacity of states, not only for epidemiologists, but also lab support,” says Thacker. For example, over 70,000 suspected anthrax samples across the United States were tested during autumn 2001, and such sudden bursts of activity put a strain on already stressed support staff.

The CSTE is playing an active role in creating on-the-job training opportunities. “We’re looking for ways to help orient people to applied epidemiology when seeking employment with state and local health departments,” says Pat McConnon, executive director of the CSTE. Together with Thacker, McConnon is developing a programme to give people who have a background in both statistics and epidemiology

## European approach

**The Epidemic Intelligence Service training programme, run by the US Centers for Disease Control and Prevention in Atlanta, Georgia, has had an impact in Europe. A group of Europeans who are former fellows of the scheme have designed the European Programme for Intervention Epidemiology Training (EPIET). Funded by the European Commission,**

**EPIET takes on 8–10 people each year and is aimed at developing the continent's capacity for responding to infectious disease. It provides up to seven years of training at the various national centres for surveillance and control of diseases in Europe.**

**Once the trainees complete the course, they are encouraged to set up similar training schemes in their home locations.**

V.G.

▶ [www.epiet.org](http://www.epiet.org)



## Local knowledge

After the US anthrax attacks in October 2001, California and Florida set up their own training programmes to improve disease surveillance.

The first class of Florida trainees is now halfway through its two-year course.

"The hardest part is keeping people in the programme," says Alan Rowan (left), who runs the scheme.

The job offers are already pouring in for the trainees — state and local public-health departments are eager to fill their vacancies with staff who have both academic and field training.

V.G.

Florida training scheme

▶ [www9.myflorida.com/disease\\_ctrl/epi/FLEIS/fleis.htm](http://www9.myflorida.com/disease_ctrl/epi/FLEIS/fleis.htm)

on-the-job training in the public-health sector. Using established state epidemiologists as mentors, this programme is designed to prepare those with a master's degree in public health or PhD-level students for future upper-level public-health positions, such as state epidemiologist.

At the moment, the scheme is focused on infectious disease, but McConnon wants to expand its scope to include areas such as chronic disease, birth defects, and occupational and environmental health. He hopes to train the first cohort of fellows this autumn, once the targeted \$1.5 million is secure. If all goes well, he anticipates training 50–100 people a year.

Such programmes will complement existing schemes that enhance specialized laboratory skills and training. The Association of Public Health Laboratories (APHL) in Washington DC currently offers one-year fellowships in laboratory training — related specifically to emerging infectious diseases — to roughly 30 postgraduates. It also offers six research fellowships in the field to qualified PhD candidates.

### IN DEMAND

Although the greatest needs are at the frontlines of disease surveillance and response, fears about bioterrorism have fuelled an explosion of interest in epidemiology, says David Savitz, an epidemiologist at the Carolina Population Center in Chapel Hill. Enrolment in the 32 US schools of public health is up and becoming more competitive.

Meir Stampfer, chair of Harvard University's department of epidemiology, acknowledges that this boost is causing problems. "We just don't have the

capacity to train more people than we are training," he says. "As a consequence, it is more difficult to gain admission here, it is more competitive."

But the career path leading from university to public health is not always clear. Undergraduate and graduate schools prepare people for the academic field, or for working in areas such as the pharmaceutical industry, says Leslie Wolf, assistant director of the North Carolina State Laboratory of Public Health in Raleigh. But she feels that many people tend to end up in public-health posts more by luck than judgement, although she notes that it is a good fit for people who gravitate towards applied research.

Although much of the current impetus for epidemiology is aimed at bioterrorism surveillance, the legacy of the new training programmes should be a workforce better prepared to deal with emerging infectious diseases, and that has gained fresh insights into chronic ailments. But Mike Osterholm, director of the University of Minnesota's Center for Infectious Disease Research and Policy in Minneapolis, suggests that 'reactionary' workforce planning may not be enough. Given the increased risk of bioterrorism and emerging infectious and chronic diseases, he feels that it is important to anticipate future needs. "We should do more forecasting and plan the workforce on what is likely to happen in the future," he says.

Virginia Gewin is a freelance science writer in Corvallis, Oregon.

CDC Epidemic Intelligence Service ▶ [www.cdc.gov/eis](http://www.cdc.gov/eis)

Council of State and Territorial Epidemiologists ▶ [www.cste.org](http://www.cste.org)

Association of Public Health Laboratories ▶ [www.aphl.org](http://www.aphl.org)

Meir Stampfer says that epidemiology courses are becoming increasingly competitive.